PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68 — A2

(11) International Publication Number: WO 00/63437

(43) International Publication Date: 26 October 2000 (26.10.00)

(21) International Application Number: PCT/US00/10716

(22) International Filing Date: 20 April 2000 (20.04.00)

(30) Priority Data:

60/130,089	20 April 1999 (20.04.99)	US
60/135,053	20 May 1999 (20.05.99)	US
60/135,051	20 May 1999 (20.05.99)	US
60/135,123	20 May 1999 (20.05.99)	US
60/161,148	22 October 1999 (22.10.99)	US
60/160,917	22 October 1999 (22.10.99)	US
60/160,927	22 October 1999 (22.10.99)	US
09/425,633	22 October 1999 (22.10.99)	US
09/513,362	25 February 2000 (25.02.00)	US
09/517,945	3 March 2000 (03.03.00)	US
09/535,854	27 March 2000 (27.03.00)	US

(71) Applicant: ILLUMINA, INC. [US/US]; Suite 200, 9390 Towne Centre Drive, San Diego, CA 92121 (US).

(72) Inventors: GUNDERSON, Kevin; 1543 Juniper Hill Drive, Encinitas, CA 92024 (US). STUELPNAGEL, John, R.; 38 Briggs Avenue, Encinitas, CA 92024 (US). CHEE, Mark, S.; 155 15th Street, #24, Del Mar, CA 92014 (US).

(74) Agents: BREZNER, David, J. et al.; Flehr Hohbach Test Albritton & Herbert LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: DETECTION OF NUCLEIC ACID REACTIONS ON BEAD ARRAYS

(57) Abstract

The present invention is directed to methods and compositions for the use of microsphere arrays to detect and quantify a number of nucleic acid reactions. The invention finds use in genotyping, i.e. the determination of the sequence of nucleic acids, particularly alterations such as nucleotide substitutions (mismatches) and single nucleotide polymorphisms (SNPs). Similarly, the invention finds use in the detection and quantification of a nucleic acid target using a variety of amplification techniques, including both signal amplification and target amplification. The methods and compositions of the invention can be used in nucleic acid sequencing reactions as well. All applications can include the use of adapter sequences to allow for universal arrays.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DETECTION OF NUCLEIC ACID REACTIONS ON BEAD ARRAYS

This application is a continuing application of: U.S.S.N.s 60/135,123, filed May 20, 1999; 60/160,917, filed October 22, 1999; 60/135,051, filed May 20, 1999; 60/161,148, filed October 22, 1999; 09/517,945, filed March 3, 2000; 60/130,089, filed April 20, 1999; 60/160,027, filed October 22, 1999; 09/513,362, filed February 25, 2000; 60/135,053, May 20, 1999; 09/425,633, filed October 22, 1999; and 09/535,854, filed March 27, 2000, all of which are expressly incorporated by reference.

FIELD OF THE INVENTION

The present invention is directed to methods and compositions for the use of microsphere arrays to detect and quantify a number of nucleic acid reactions. The invention finds use in genotyping, i.e. the determination of the sequence of nucleic acids, particularly alterations such as nucleotide substitutions (mismatches) and single nucleotide polymorphisms (SNPs). Similarly, the invention finds use in the detection and quantification of a nucleic acid target using a variety of amplification techniques, including both signal amplification and target amplification. The methods and compositions of the invention can be used in nucleic acid sequencing reactions as well. All applications can include the use of adapter sequences to allow for universal arrays.

BACKGROUND OF THE INVENTION

The detection of specific nucleic acids is an important tool for diagnostic medicine and molecular biology research. Gene probe assays currently play roles in identifying infectious organisms such as bacteria and viruses, in probing the expression of normal and mutant genes and identifying mutant genes such as oncogenes, in typing tissue for compatibility preceding tissue transplantation, in matching tissue or blood samples for forensic medicine, and for exploring homology among genes from different species.

Ideally, a gene probe assay should be sensitive, specific and easily automatable (for a review, see Nickerson, Current Opinion in Biotechnology 4:48-51 (1993)). The requirement for sensitivity (i.e. low detection limits) has been greatly alleviated by the development of the polymerase chain reaction

(PCR) end other amplification technologies which ellow reseerchars to emplify exponentially a specific nucleic ecid sequence before analysis (for a review, see Abramson et al., Current Opinion In Biotechnology, 4:41-47 (1993)).

Sensitivity, i.e. detection limits, remain a significant obstacle in nucleic acid detection systems, and a variety of techniquas have been developed to address this issue. Briafly, these techniques can be classified as either target emplificetion or signal amplification. Target amplification involves the amplification (i.e. replication) of the target sequence to be detected, resulting in a significant increase in tha number of target molecules. Target amplification strategies includa the polymarase chain reaction (PCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA).

Altemalively, rather than amplify the target, altamate techniques use the target as a template to replicata a signalling probe, allowing a small number of target molecules to result in a large number of signelling probes, that then can ba detected. Signal amplification strategies include the ligase chain raaction (LCR), cycling probe technology (CPT), invasive cleavage techniques such as Invader™ technology, Q-Beta replicase (QβR) technology, and the usa of "amplification probes" such as "brenched DNA" that result in multiple label probes binding to a single target sequence.

The polymerase chain reaction (PCR) is widely used end described, and involves the use of primer extension combined with thermal cycling to amplify a target sequence; sea U.S. Patent Nos. 4,683,195 and 4,683,202, and PCR Essential Data, J. W. Wiley & sons, Ed. C.R. Newton, 1995, all of which are incorporated by reference. In addition, there are a number of variations of PCR which also find use in the invention, including "quantitative competitive PCR" or "QC-PCR", "arbitrarily primed PCR" or "AP-PCR" , "immuno-PCR", "Alu-PCR", "PCR singla strand conformational polymorphism" or "PCR-SSCP", allelic PCR (see Newton et al. Nucl. Acid Res. 17:2503 91989); "reverse transcriptasa PCR" or "RT-PCR", "biotin capture PCR", "vectorette PCR". "panhandle PCR", and "PCR eelect cDNA subtraction", emong others.

Strand displacement amplification (SDA) is generally described in Walker et al., in Molecular Methods for Virus Detection, Academic Press, Inc., 1995, end U.S. Patent Nos. 5,455,168 end 5,130,238, all of which ere hereby incorporated by reference.

Nucleic ecid sequence based amplification (NASBA) is generally described in U.S. Patent No. 5,409,818 and "Profiting from Gene-based Diagnostics", CTB International Publishing Inc., N.J., 1996, both of which ere incorporated by reference.

Cycling probe technology (CPT) is a nucleic ecid detection system based on signal or probe

amplification rather than target amplification, such es is done in polymerase chain reactions (PCR). Cycling probe technology relies on a molar excess of labeled probe which contains a scissile linkage of RNA. Upon hybridization of the probe to tha target, the resulting hybrid contains e portion of RNA:DNA. This area of RNA:DNA duplex is recognized by RNAseH end the RNA is excised, resulting in cleavage of the probe. The probe now consists of two smaller sequences which may be released, thus leaving the target intact for repeated rounds of the reaction. The unreacted probe is removed and the label is then detected. CPT is generally described in U.S. Patent Nos. 5,011,769, 5,403,711, 5,660,988, and 4,876,187, and PCT published applications WO 95/05480, WO 95/1416, and WO 95/00667, all of which are specifically incorporated herein by reference.

The oligonucleotide ligation assay (OLA; sometimes referred to es the ligetion chain reaction (LCR)) involve the ligation of at least two smaller probes into a single long probe, using the target sequence as the template for the ligase. See generally U.S. Patent Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; and WO 89/09835, all of whlch are incorporated by reference.

Invader™ technology is based on structure-specific polymereses that cleave nucleic ecids in e site-specific manner. Two probes are used: en "invader" probe and a "signalling" probe, that adjacently hybridize to a target sequence with a non-complementary overlap. The enzyme cleaves at the overlap due to its recognition of the "tail", and releases the "tail" with a label. This cen then be detected. The Invader™ technology is described in U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; end 5,843,669, all of which are hereby incorporated by reference.

"Rolling circle amplification" is based on extension of a circular probe that has hybridized to a target sequence. A polymerase is added that extends the probe sequence. As the circular probe has no terminus, the polymerase repeatedly extends the circular probe resulting in concatamers of the circuler probe. As such, the probe is amplified. Rolling-circle amplification is generally described in Baner *et al.* (1998) *Nuc. Acids Res.* 26:5073-5078; Barany, F. (1991) *Proc. Natl. Aced. Sci. USA* 88:189-193; and Lizardi *et al.* (1998) *Nat. Genet.* 19:225-232, all of which are incorporated by reference in their entirety.

"Branched DNA" signal amplification relies on the synthesis of branched nucleic ecids, containing a multiplicity of nucleic acid "arms" that function to increase the amount of label that can be put onto one probe. This technology is generally described in U.S. Patent Nos. 5,681,702, 5,597,909, 5,545,730, 5,594,117, 5,591,584, 5,571,670, 5,580,731, 5,571,670, 5,591,584, 5,624,802, 5,635,352, 5,594,118, 5,359,100, 5,124,246 and 5,681,697, all of which are hereby incorporated by reference.

Similarly, dendrimers of nucleic acids serve to vastly increase the amount of label that can be added to a single molecule, using a similar idea but different compositions. This technology is as described

In U.S. Patent No. 5,175,270 and Nilsen et al., J. Theor. Biol. 187:273 (1997), both of which ere incorporeted herein by reference.

Specificity, In contrast, remains a problem in many currently available gene probe esseys. The extent of molecuier complementarity between probe end target defines the specificity of the interaction. In a practical sense, the degree of similarity between the target end other sequences In the sample also has an Impact on specificity. Variations in the concentrations of probes, of targets end of salts In the hybridizetion medium, In the reaction tempereture, end in the length of the probe may elter or Influence the specificity of the probe/target Interaction.

It may be possible under some circumstances to distinguish targets with perfect complementarity from targets with mismatches; this is generelly very difficult using traditionel technology such es filter hybridizetion, In situ hybridization etc., since small veriations In the reection conditions will elter the hybridization, elthough this may not be a problem if appropriate mismatch controls are provided. New experimental techniques for mismatch detection with standard probes include DNA ligation asseys where single point mismatches prevent ligation end probe digestion essays in which mismetches create sites for probe cleavege.

Recent focus hes been on the enelysis of the reletionship between genetic varietion end phenotype by making use of polymorphic DNA markers. Previous work utilized short tandem repeats (STRs) as polymorphic positional markers; however, recent focus is on the use of single nucleotide polymorphisms (SNPs), which occur et en everage frequency of more than 1 per kilobase In humen genomic DNA. Some SNPs, perticularly those In and eround coding sequences, ere likely to be the direct cause of therapeutically relevent phenotypic variants end/or disease predisposition. There are a number of well known polymorphisms that cause clinically important phenotypes; for example, the epoE2/3/4 variants ere essociated with different relative risk of Alzheimer's end other diseases (see Cordor et al., Science 261(1993). Multiplex PCR amplification of SNP loci with subsequent hybridization to oligonucleotide arrays has been shown to be an accurate and reliable method of simultaneously genotyping at least hundreds of SNPs; see Wang et el., Science, 280:1077 (1998); see elso Schafer et el., Neture Biotechnology 16:33-39 (1998). The compositions of the present Invention mey easily be substituted for the arrays of the prior ert.

There are a variety of particular techniques thet ere used to detect sequence, including mutetions and SNPs. These include, but are not limited to, ligation based essays, cleavage based essays (mismatch end Invasive cleevege such es Invader™), single base extension methods (see WO 92/15712, EP 0 371 437 B1, EP 0317 074 B1; Pastinen et el., Genome Res. 7:606-614 (1997); Syvänen, Clinica Chimica Acta 226:225-236 (1994); and WO 91/13075), and competitive probe analysis (e.g. competitive sequencing by hybridization; see below).

In addition, DNA sequencing is e cruciel technology in biology todey, es the repid eequencing of genomes, Inciuding tha human genoma, is both e significant goel and e significant hurdle. Thus thare is e significent need for robust, high-throughput methods. Traditionally, the most common method of DNA sequencing has been based on polyacrylemide gel frectionetion to resolve a populetion of chein-terminated fragments (Sangar et ei., Proc. Netl. Acad. Sci. USA 74:5463 (1977); Maxam & Gilbert). The populetion of fragments, termineted at each position in the DNA sequence, can be generated in e number of ways. Typicaily, DNA polymerase is used to incorporate didaoxynucleotides that serve es chain terminators.

Several eltemative methods have been developed to increase the speed end eese of DNA sequencing. For example, sequencing by hybridization has been described (Drmenec et el., Genomics 4:114 (1989); Koster et al., Neture Biotachnology 14:1123 (1996); U.S. Patent Nos. 5,525,464; 5,202,231 end 5,695,940, emong others). Similerly, sequencing by synthesis is an altemetive to gel-based sequencing. These methods add end read only one base (or et most e few beses, typically of tha same type) prior to polymerization of the next bese. This can be referred to as "time resolved" sequencing, to contrast from "gel-resolved" sequencing. Sequencing by synthesis has been described in U. S. Patent No 4,971,903 end Hyman, Anal. Biochem. 174:423 (1988); Rosenthal, International Patent Application Publication 761107 (1989); Metzker et ei., Nucl. Acids Res. 22:4259 (1994); Jones, Biotechniquas 22:938 (1997); Ronaghi et al., Anal. Biochem. 242:84 (1996), Nyren et ei., Anal. Biochem. 151:504 (1985). Detection of ATP sulfurylase activity is described in Karamohamed end Nyren, Anal. Biochem. 271:81 (1999). Sequencing using reversible chain terminating nucleotides is dascribed in U.S. Patent Nos. 5,902,723 and 5,547,839, and Canard and Arzumanov, Gene 11:1 (1994), and Dyatkina and Arzumanov, Nucleic Acids Symp Ser 18:117 (1987). Reversible chein termination with DNA ligase is described in U.S. Patent 5,403,708. Time resolvad saquencing is described in Johnson at ai., Anal. Biocham. 136:192 (1984). Singie molacule analysis is described in U.S. Patent No. 5,795,782 end Elgen and Rigler, Proc. Natl Acad Scl USA 91(13):5740 (1994), all of which ere hereby expressly incorporeted by reference in their entirety.

One promising sequencing by synthesis method is based on the detection of the pyrophosphate (PPi) released during the DNA polymeresa reaction. As nucleotriphosphates ere edded to e growing nucleic ecid chain, they reieese PPi. This release can be quentitatively measured by the conversion of PPi to ATP by the enzyme sulfurylase, and the subsequent production of visible light by firefly luciferase.

Several essay systems heve been described that capitalize on this mechanism. See for exempla WO93/23564, WO 98/28440 end WO98/13523, ell of which are expressly incorporated by reference. A preferred method is described in Ronaghi et el., Science 281:363 (1998). In this method, the four deoxynucleotides (dATP, dGTP, dCTP end dTTP; collectively dNTPs) ere edded stepwise to e pertial duplex comprising e saquencing primer hybridized to a single stranded DNA template and incubated with DNA polymerese, ATP sulfurylese, lucifarase, and optionaiiy a nucieotide-degrading enzyme such

as apyrase. A dNTP is only incorporated into the growing DNA strand if it is complementary to the base in the template strend. The synthesis of DNA is accompanied by the releese of PPi equal in molarity to the incorporated dNTP. The PPi is converted to ATP and the light generated by the luciferase is directly proportional to the amount of ATP. In some cases the unincorporated dNTPs and the produced ATP are degraded between each cycle by the nucleotide degrading enzyme.

In some cases the DNA template is associated with e solid support. To this end, there are a wide variety of known methods of attaching DNAs to solid supports. Recent work has focused on the attachment of binding ligands, including nucleic acid probes, to microspheres that are randomly distributed on a surface, including e fiber optic bundle, to form high density arrays. See for example PCTs US98/21193, PCT US99/14387 end PCT US98/05025; WO98/50782; and U.S.S.N.s 09/287,573, 09/151,877, 09/256,943, 09/316,154, 60/119,323, 09/315,584; all of which are expressly incorporated by reference.

An additional technique utilizes sequencing by hybridization. For example, sequencing by hybridization has been described (Drmanac et el., Genomics 4:114 (1989); U.S. Patent Nos. 5,525,464; 5,202,231 and 5,695,940, among others, all of which are hereby expressly incorporated by reference in their entirety).

In addition, sequencing using mass spectrometry techniques has been described; see Koster et al., Nature Biotechnology 14:1123 (1996).

Finally, the use of adapter-type sequences that allow the use of universal arrays has been described in limited contexts; see for example Chee et el., Nucl. Acid Res. 19:3301 (1991); Shoemaker et al., Nature Genetics 14:450 (1998); Barany, F. (1991) Proc. Netl. Aced. Sci. USA 88:189-193; EP 0 799 897 A1; WO 97/31256, all of which are expressly incorporated by reference.

PCTs US98/21193, PCT US99/14387 and PCT US98/05025; WO98/50782; and U.S.S.N.s 09/287,573, 09/151,877, 09/256,943, 09/316,154, 60/119,323, 09/315,584; all of which are expressly incorporated by reference, describe novel compositions utilizing substrates with microsphere arrays, which allow for novel detection methods of nucleic ecid hybridization.

Accordingly, it is an object of the present invention to provide detection and quantification methods for a variety of nucleic acid reactions, including genotyping, amplification reactions and sequencing reactions, utilizing microsphere arrays.

SUMMARY OF THE INVENTION

In accordance with the above objects, the present invention provides methods of determining the

identity of a nucleotida at e detaction position in a target saquenca. Tha mathods comprisa providing a hybridization complex comprising the targat eequence end e captura probe covelently attachad to e microsphere on a surface of a substrate. The mathods comprisa datarmining tha nuclaotida at tha detection position. The hybridization complax can comprisa the captura proba, a capture extandar probe, and tha targat sequance. In addition, tha target sequance may comprisa axogeneous adapter sequencas.

In en additional aspact, the method comprises contacting tha microspharas with a plurality of detection probes each comprising a unique nucleotide at the reedout position end e unique datactable label. Tha signal from at least one of the detectable labels is detacted to idantify tha nuclaotida et the detection position.

In an edditional aspect, the datection probe does not contain detection label, but rather is idantified based on its charactaristic mass, for exampla via mass spectromatry. In addition, the detection proba comprises a unique label that is detectad based on its characteristic mass.

In a furthar aspect, the invention providas methods wherein tha targat sequance comprises a first targat domain directly 5' adjacant to the detection position. The hybridization complex comprises tha target sequence, a capture probe and an axtension primer hybridized to tha first target domain of tha target sequanca. The datermination step comprises contacting the microsphares with e polymerase enzyme, and a plurality of NTPs each comprising a covalently attached detactabla label, under conditions whereby if ona of the NTPs basepairs with the base at the detection position, tha extension primer is extended by the enzyme to incorporate tha label. As is known to thosa in the art, dNTPs and ddNTPs are tha prefarred substrates for DNA polymerases. NTPs ara the prafarred substrates for RNA polymerasas. Tha base et the datection position is then idantifiad.

In an additional aspect, tha invention provides mathods whereln the target sequance comprises a first target domain directly 5' adjacent to the datection position, wherain the captura probe serves as an extenslon primer and is hybridizad to the first targat domain of the target sequance. Tha determination step comprises contacting tha microspheras with a polymerase enzyme, end a plurality of NTPs each comprising a covalently ettached datectabla label, under conditions whereby if one of the NTPs basepairs with the base et tha detaction position, the axtension primar is axtended by the anzyme to incorporata tha label. Tha base at tha datection position is thus identifiad.

In a further espact, tha invention providas methods wherein tha target sequence comprises (5' to 3'), a first targat domain comprising en overlap domain comprising et least a nucleotide in the detection position and a second targat domain contiguous with the datection position. Tha hybridization complax comprises a first probe hybridized to tha first target domain, and e second probe hybridized to tha sacond targat domain. The sacond probe comprises a dataction sequenca that does not hybridize

with the target sequence, end a detectable lebel. If the second probe comprises a base that is perfectly complementary to the detection position e cleavage structure is formed. The method further comprises contacting the hybridization complex with e cleavage enzyme that will cleave the detection sequence from the signelling probe and then forming an assay complex with the detection sequence, a capture probe covalently attached to a microsphere on a surfece of a substrate, end et leest one lebel. The base at the detection position is thus identified.

In en additional espect, the invention provides methods of determining the identification of a nucleotide at a detection position in e target sequence comprising a first target domain comprising the detection position end a second target domain edjacent to the detection position. The method comprises hybridizing a first ligetion probe to the first target domein, and hybridizing a second ligation probe to the second target domain. If the second ligation probe comprises a base that is perfectly complementary to the detection position a ligation structure is formed. A ligation enzyme is provided that will ligate the first end the second ligation probes to form e ligeted probe. An essay complex is formed with the ligated probe, a capture probe covalently attached to e microsphere on e surface of a substrate, and at least one label. The base at the detection position is thus identified.

In a further aspect, the present invention provides methods of sequencing e plurality of target nucleic acids. The methods comprise providing a plurality of hybridization complexes each comprising a target sequence end a sequencing primer that hybridizes to the first domain of the target sequence, the hybridization complexes ere ettached to e surfece of a substrate. The methods comprise extending each of the primers by the eddition of a first nucleotide to the first detection position using an enzyme to form en extended primer. The methods comprise detecting the release of pyrophosphete (PPi) to determine the type of the first nucleotide edded onto the primers. In one espect the hybridizetion complexes are attached to microspheres distributed on the surface. In an additional aspect the sequencing primers ere attached to the surface. The hybridizetion complexes comprise the target sequence, the sequencing primer and a capture probe covalently attached to the surface. The hybridization complexes also comprise an edapter probe.

In en additionel espect, the method comprises extending the extended primer by the addition of a second nucleotide to the second detection position using an enzyme and detecting the release of pyrophosphate to determine the type of second nucleotide edded onto the primers. In en additionel aspect, the pyrophosphate is detected by contacting the pyrophosphete with e second enzyme that converts pyrophosphate into ATP, and detecting the ATP using a third enzyme. In one aspect, the second enzyme is sulfuryiase end/or the third enzyme is luciferase.

In en additional espect, the invention provides methods of sequencing e target nucleic acid comprising e first domain end an edjacent second domein, the second domain comprising e plurality of terget positions. The method comprises providing e hybridization complex comprising the terget sequence

and a capture probe covalently attached to microspheres on a surface of a substrate and determining the identity of a plurality of bases at the target positions. The hybridization complex comprises the capture probe, an adapter probe, and the target sequence. In one aspect the sequencing primer is the capture probe.

In an additional aspect of the invention, the determining comprises providing a sequencing primer hybridized to the second domain, extending the primer by the addition of first nucleotide to the first detection position using a first enzyme to form an extended primer, detecting the release of pyrophosphate to determine the type of the first nucleotide added onto the primer, extending the primer by the addition of a second nucleotide to the second detection position using the enzyme, and detecting the release of pyrophosphate to determine the type of the second nucleotide added onto the primer. In an additional aspect pyrophosphate is detected by contacting the pyrophosphate with the second enzyme that converts pyrophosphate into ATP, and detecting the ATP using a third enzyme. In one aspect the second enzyme is sulfurylase and/or the third enzyme is luciferase.

In an additional aspect of the method for sequencing, the determining comprises providing a sequencing primer hybridized to the second domain, extending the primer by the addition of a first protected nucleotide using a first enzyme to form an extended primer, determining the identification of the first protected nucleotide, removing the protection group, adding a second protected nucleotide using the enzyme, and determining the identification of the second protected nucleotide.

In an additional aspect the invention provides a kit for nucleic acid sequencing comprising a composition comprising a substrate with a surface comprising discrete sites and a population of microspheres distributed on the sites, wherein the microspheres comprise capture probes. The kit also comprises an extension enzyme and dNTPs. The kit also comprises a second enzyme for the conversion of pyrophosphate to ATP and a third enzyme for the detection of ATP. In one aspect the dNTPs are labeled. In addition each dNTP comprises a different label.

In a further aspect, the present invention provides methods of detecting a target nucleic acid sequence comprising attaching a first adapter nucleic acid to a first target nucleic acid sequence to form a modified first target nucleic acid sequence, and contacting the modified first target nucleic acid sequence with an array as outlined herein. The presence of the modified first target nucleic acid sequence is then detected.

In an additional aspect, the methods further comprise attaching a second adapter nucleic acid to a second target nucleic acid sequence to form a modified second target nucleic acid sequence and contacting the modified second target nucleic acid sequence with the array.

In a further aspect, the invention provides methods of detecting a target nucleic acid sequence

comprising hybridizing a first primer to a first portion of a target sequence, wherain the first primer further comprises en adapter sequenca end hybridizing e second primar to a second portion of tha targel sequence. Tha first and sacond primers are ligated together to form a modlfiad primer, end the adapter sequence of tha modified primer is contacted with en array of tha invention, to allow detection of tha presenca of tha modified primer.

In an additional ambodimant, the prasant Invantion provides e mathod for datecting a first targat nucleic acid sequance. In ona aspect tha mathod comprises hybridizing at least a first primer nucleic acid to tha first target sequenca to form a first hybridization complax, contacting tha first hybridization complex with a first anzyme to form a modifiad first primer nuclaic ecid, disessociating the first hybridizetion complex, contacting the modified first primer nucleic acid with an array comprising a substrata with a surface comprising discrata sitas and a population of microspheres comprising at least a first subpopulation comprising a first cepture proba such that tha first captura probe end the modified primar form en assay complex, wherein the microsphares ere distributad on tha surfaca, and datacting the presenca of the modified primer nuclaic acid.

In eddition the method further comprisas hybridizing at least a second primer nucleic acid to a second target sequence thet is substantially complementary to tha first target sequenca to form a second hybridization complax, contacting the second hybridization complex with the first enzyme to form modified sacond primer nucleic ecid, disassociating the second hybridization complex and forming a second assay complex comprising the modified second primer nucleic acid and a second capture probe on a second subpopulation.

In an additional aspect of the invantion tha primar forms a circular probe following hybridization with the target nuclaic acid to form a first hybridization complax and contacting the first hybridization complex with e first enzyme comprising e ligase such thet tha oligonucleotide ligation essay (OLA) occurs. This is followed by adding tha second enzyme, a polymerase, such that the circular proba is amplified in e rolling circla amplification (RCA) assay.

In an additional espect of the invention, the first enzyme comprises e DNA polymerese and the modification is an extension of the primer such that the polymerase chein reaction (PCR) occurs. In an additional aspact of the Invantion the first anzyme comprises a ligase and the modification comprises a ligation of the first primer whlch hybridizes to a first domain of the first target sequence, to a third primar which hybridizes to e second edjacant domain of the first targat saquenca such that the ligese chain raaction (LCR) occurs.

In an additional aspect of the invention, the first primer comprises e first probe sequence, a first scissila linkage end e second proba sequence, wherein the first enzyme will claava tha scissila linkaga resulting in the separation of the first and second probe sequances and tha disassociation of the first

hybridizetion complex, leaving the first target sequence intact such that the cycling probe technology (CPT) reaction occurs.

In addition, wherein the first enzyme is a polymerase that extends the first primer and the modified first primer comprises e first newly synthesized strand, the method further comprises the eddition of e second enzyme comprising e nicking enzyme that nicks the extended first primer leaving the first target sequence intact, end extending from the nick using the polymerase, and thereby displacing the first newly synthesized strand end generating a second newly synthesized strend such thet strand displacement emplification (SBA) occurs.

In addition, wherein the first target sequence is an RNA target sequence, the first primer nucleic ecid is e DNA primer comprising en RNA polymerase promoter, the first enzyme is a reverse-transcriptase that extends the first primer to form e first newly synthesized DNA strand, the method further comprises the eddition of e second enzyme comprising an RNA degreding enzyme that degrades the first target sequence, the addition of a third primer that hybridizes to the first newly synthesized DNA strand, the addition of a third enzyme comprising a DNA polymerase thet extends the third primer to form a second newly synthesized DNA strand, to form a newly synthesized DNA hybrid, the addition of a fourth enzyme comprising en RNA polymerese that recognizes the RNA polymerase promoter and generates at leest one newly synthesized RNA strand from the DNA hybrid, such that nucleic ecid sequence-based amplification (NASBA) occurs.

In addition, wherein the first primer is en inveder primer, the method further comprises hybridizing a signalling primer to the target sequence, the enzyme comprises a structure-specific cleeving enzyme and the modification comprises a cleavege of sald signalling primer, such that the invasive cleevege reaction occurs.

An additional aspect of the invention is e method for detecting a target nucleic acid sequence comprising hybridizing e first primer to e first target sequence to form e first hybridization complex, contacting the first hybridization complex with e first enzyme to extend the first primer to form a first newly synthesized strand and form e nucleic acid hybrid thet comprises en RNA polymerase promoter, contacting the hybrid with en RNA polymerase that recognizes the RNA polymerase promoter end generates at least one newly synthesized RNA strand, contacting the newly synthesized RNA strand with en erray comprising e substrate with e surface comprising discrete sites end e populetion of microspheres comprising et leest a first subpopulation comprising e first capture probe; such that the first capture probe end the modified primer form an essay complex; wherein the microspheres ere distributed on the surface end detecting the presence of the newly synthesized RNA strand.

In addition, when the target nucleic ecid sequence is an RNA sequence, end prior to hybridizing a first primer to a first target sequence to form a first hybridization complex, method comprises hybridizing a

sacond primer comprising an RNA polymarasa promotar sequanca to tha RNA saquance to form a sacond hybridization complex, contacting the second hybridization complex with a second enzyme to extand the second primar to form e sacond nawly synthasized strand end form a nuclaic ecid hybrid; and dagrading tha RNA sequence to leave the second newly synthesized strend as the first target sequanca. In ona aspact of tha invention tha degrading is dona by the addition of an RNA dagrading enzyma. In an additional aspact of tha invention the degreding is done by RNA degrading activity of reversa transcriptasa.

In addition, when the targat nuclaic acid sequance is e DNA saquenca, and prior to hybridizing e first primer to e first targat sequence to form a first hybridization complex, tha mathod comprisas hybridizing a second primer comprising an RNA polymarase promoter sequanca to tha DNA sequanca to form a sacond hybridizetion complax, contacting tha second hybridization complax with a sacond anzyme to extend the second primer to form a sacond newly synthasized strand end form a nucleic acid hybrid, and danaturing the nucleic acid hybrid such that tha second newly synthesizad strand is the first target sequence.

An additional aspect fo the invantion is a kit for the detection of a first target nucleic acid sequenca. Tha kit comprises at least a first nucleic acid primer substantially complemantary to at laast a first domain of tha target sequence, at least a first enzyma that will modify tha first nuclaic acid primer, and en array comprising a substrate with e surface comprising discrete sites, and e population of microspheres comprising at laest a first and a second subpopulation, wharein each subpopulation comprises a bioactive egent, wherein the microsphares are distributed on tha surface.

In an additional aspect of the invention, is a kit for the detection of a PCR reaction wherein the first enzyme is a thermostabla DNA polymerasa.

In an additional aspect of tha invention, is a kit for the detection of a LCR reaction wherein tha first enzyme is a ligese and the kit comprises a first nucleic acid primer substantially complemantary to e first domain of tha first target sequenca and a third nucleic ecid primer substantially complamentary to e second adjacent domain of tha first terget sequence.

In en edditional espect of the invention, is e kit for the detection of e strand displecement amplification (SDA) raaction wherein the first anzyma is e polymerase and tha kit further comprises a nicking anzyme.

In en additional aspect of the invention, is a kit for tha detection of a NASBA reaction wherein the first enzyme is a reverse transcriptase, and tha kit comprises a second anzyme comprising an RNA degrading enzyme, a third primer, a third enzyme comprising a DNA polymerasa and a fourth enzyme comprising an RNA polymarasa.

In an additional espect of the invention, is a kit for the detection of en invasive cleavage reaction wharein tha first anzyma is a structure-specific claaving enzyme, and tha kit comprises a signaling primer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A, 1B end 1C depict three different embodiments for etteching a target sequence to an array. The solid support **5** has microsphere **10** with captura probe **20** linked via a linker **15**. Figura 1A dapicts diract attachmant; tha captura proba **20** hybridizes to a first portion of the targat sequance **25**. Figure 1B depicts the use of a capture extender probe **30** that hes e first portion that hybridizes to the captura proba **20** and a sacond portion that hybridizes to a first domein of tha targat sequanca **25**. Figura 1C shows the use of an adaptar sequanca **35**, that has been edded to tha target saquanca, for example during an amplification reaction as outlined herein.

Figuras 2A and 2B depict two preferrad embodiments of SBE amplification. Figura 2A shows axtension primar **40** hybridized to the target saquance **25**. Upon eddition of the extension enzyme and labelled nuclaotides, the extension primer is modified to form a labelled primar **41**. The reaction can ba repeated end then the labelled primar is added to tha array as above. Figura 2B dapicts the same reaction but using adapter saquences.

Figures 3A end 3B dapict two preferrad embodiments of OLA emplification. Figura 3A depicts a first ligation proba **45** and e second ligation probe **50** with a label **55**. Upon addition of the ligase, the probes are ligated. The reaction cen be repeated and then tha ligatad primer is added to the errey as above. Figure 3B dapicts tha sama reaction but using edaptar sequencas.

Figura **4** depicts a praferrad embodiment of tha invasive cleavege reaction. In this embodiment, the signaling proba **65** comprisas two portions, e detection saquence **67** end e signaling portion **66**. Tha signaling portion can sarva as an adapter sequenca. In addition, the signaling portion generally comprises the labal **55**, although as will be appreciated by those in tha art, the label may be on the detection saquence as well. In eddition, for optional ramovel of tha unclaaved probes, a capture tag **60** may elso be used. Upon eddition of the enzyme, tha structure is cleaved, releasing tha signaling portion **66**. The raaction cen be rapeated and than the signaling portion is added to the array as ebove.

Figures 5A and 5B dapict two prefarred ambodimants of CPT amplification. A CPT primer **70** comprising e labal **55**, a first probe sequenca **71** end a second probe sequence **73**, saparatad by a scissile linkaga **72**, and optionally comprising a captura tag **60**, is hybridized to tha terget sequence **25**. Upon addition of the enzyma, tha scissile linkage is cleaved. The reaction can ba rapeated and then the proba sequance comprising tha label is added to the array as above. Figura 5B depicts tha same

reection but using edepter sequences.

Figure 6 depicts OLA/RCA emplification using e single "padlock probe" 57. The padlock probe is hybridized with e target sequence 25. When the probe 57 is complementary to the target sequence 26, ligation of the probe terminl occurs forming e circular probe 28. When the probe 57 ie not complementary to the target sequence 27, ligation does not occur. Addition of polymerase and nucleotides to the circuler probe results amplificetion of the probe 58. Cleavage of the emplified probe 58 yields fregments 59 that hybridize with an identifier probe 21 immobilized on a microsphere 10.

Figure 7 depicts en elternetive method of OLA/RCA. An immobilized first OLA primer 45 is hybridized with e target sequence 25 end e second OLA primer 50. Following the eddition of ligase, the first and second OLA primers ere ligated to form e ligated oligonucleotide 56. Following denaturation to remove the target nucleic acid, the immobilized ligeted oligonucleotide is distributed on en erray. An RCA probe 57 end polymerase are added to the array resulting in amplificetion of the circular RCA probe 58.

Figures 8A, 8B, 8C, 8D and 8E schematically depict the use of readout probes for genotyping. Figure 8A shows a "sandwich" format. Substrete 5 hes e discrete site with e microsphere 10 comprising e capture probe 20 ettached via a linker 15. The target sequence 25 hes e first domain that hybridizes to the capture probe 20 and a second domain comprising e detection position 30 that hybridizes to e readout probe 40 with readout position 35. As will be eppreciated by those in the ert, Figure 8A depicts a single detection position; however, depending on the system, e plurality of different probes can hybridize to different target domains; hence n is en integer of 1 or greater. Figure 8B depicts the use of e capture probe 20 that elso serves es a readout probe. Figure 8C depicts the use of en adapter probe 100 that binds to both the capture probe 20 and the target sequence 25. As will be appreciated by those in the ert, the figure depicts that the capture probe 20 and target sequence 25 bind edjacently end as such may be ligated; however, es will be appreciated by those in the ert, there mey be e "gep" of one or more nucleotides. Figure 8D depicts a solution based assey. Two readout probes 40, eech with a different readout position (35 end 36) and different labels (45 end 46) ere added to target sequence 25 with detection position 35, to form e hybridization complex with the match probe. This is edded to the array; Figure 8D depicts the use of e capture probe 20 that directly hybridizes to a first domein of the target sequence, elthough other ettachments may be done. Figure 8E depicts the direct ettachment of the target sequence to the array.

Figures 9A, 9B, 9C, 9D, 9E, 9F and 9G depict preferred embodiments for SBE genotyping. Figure 9A depicts a "sandwich" essay, in which substrate 5 hes e discrete site with a microsphere 10 comprising e cepture probe 20 ettached via e linker 15. The target sequence 25 hes e first domain that hybridizes

to the cepture probe 20 end a second domeln comprising a detection position 30 that hybridizes to en extension primer 50. As will be eppreciated by those In the ert, Figure 9A depicts a singla detection position; however, depending on the system, e plurality of different primers can hybridize to different target domeins; hence n is en integer of 1 or greater. In addition, the first domain of the target sequence may be en edapter sequence. Figure 9B depicts the use of a capture probe 20 that also serves as an extension primer. Figure 9C depicts the solution reaction. Figure 9D depicts the use of a capture extender probe 100, that hes a first domain thet will hybridize to the cepture probe 20 and a second domeln that will hybridize to a first domain of the target sequence 25. Figure 9E depicts the eddition of a ligation step prior to extension of the extension probe. Figure 9F depicts the addition of a ligetion step after the extension of the extenslon probe. Figure 9G deplcts the SBE solution reaction followed by hybridization of the product of tha reaction to the bead array to capture an adapter sequence.

Figures 10A, 10B, 10C, 10D and 10E depict some of the OLA genotyping embodiments of the reaction. Figure 10A depicts the solution reaction, wherein the target sequence 25 with a detection position 30 hybridizes to the first ligation probe 75 with readout position 35 and second probe 76 with a detectable label 45. As will be appreciated by those in the art, the second ligation probe could also contain the readout position. The addition of e ligase forms a ligated probe 80, that can then be added to the array with a capture probe 20. Figure 10B depicts an "on bead" assay, wherein the capture probe 20 serves as the first ligation probe. Figure 10C depicts a sandwich assay, using a capture probe 20 that hybridizes to e first portion of the target sequence 25 (which may be an endogeneous sequence or en exogeneous adapter sequence) and ligation probes 75 and 76 thet hybridize to e second portion of the target sequence comprising the detection position 30. Figure 10D depects the use of a capture extender probe 100. Figure 10E depicts a solution based assay with the use of an adapter sequence 110.

Figures 11A, 11B end 11C deplct the SPOLA reaction for genotyping. In Figure 11A, two ligation probes are hybridized to e target sequence. As will be appreciated by those in the art, this system requires that the two ligation probes be attached at different ends, i.e. one at the 5' end and one at the 3' end. One of the ligation probes is attached via a cleavable linker. Upon formation of the assay complex and tha eddition of e ligase, the two probes will efficiently covalently couple the two ligation probes if perfect complementarity at tha junction exists. In Figure 11B, the resulting ligation difference between correctly matched probes and imperfect probes is shown. Figure 11C shows that subsequent cleavage of the cleavable linker produces a reactive group, in this case an emine, that may be subsequently lebeled as outlined herein. Alternatively, cleavage may leave an upstream oligo with a detecteble lebel. If not ligated, this labaled oligo can be washed eway.

Figures 12A and 12B depict two cleavage reactions for genotyping. Figure 12A depicts a loss of signal assay, wherein a lebel 45 is cleaved off due to the discrimination of the cleavage enzyme such

as a restriction endonucleese or resolvase type enzyme to allow singla basa mismetch discrimination. Figure 12B depicts the use of e quencher **46**.

Figure 13A, 13B, 13C, 13D, 13E end 13F deplct the use of Invesive cleevege to determine the identity of the nucleotide at the detection position. Figures 13A and 13B depict e loss of signal assay. Figure 13A depicts the invader probe **55** with readout position **35** hybridized to the target sequence **25** which is ettached via a cepture probe **20** to the surfece. The signal probe **60** with readout position **35**, detectable label **45** and detection sequence **65** also binds to the target sequence **25**; the two probes form a cleavage structure. If the two reedout positions 35are capable of besepeiring to tha detection position **30** the eddition of e structure-specific cleavage enzyma releases the detection sequence **65** end consequently the lebel **45**, leading to e loss of signal. Figure 13B is the same, except that the capture probe **20** elso serves es the invader probe. Figure 13C depicts e solution reaction, wherein the signalling probe can comprise e capture tag **70** to facilitate the removal of uncleeved signel probes. The addition of the cleeved signal probe (e.g. the detection sequence **65**) with its associated label **45** results in detection. Figure 13D depicts e solution based essey using a label probe **120**. Figure 13E depicts a preferred embodiment of an invesive cleavage reaction that utilizes a fluorophore-quencher reaction. Figure 13E has the 3' end of the signal probe **60** is ettached to the bead **10** and comprises e label **45** and a quencher **46**. Upon formetion of the assay complex and subsequent cleavage, tha quencher **46** is removed, leaving the fluorophore **45**.

Figures 14A, 14B, 14C and 14D depict genotyping assays besed on the novel combination of competitive hybridizetion and extension. Figures 14A, 14B end 14C deplct solution besed asseys. After hybridization of the extension probe **50** with a match base at the readout position **35**, en extension enzyme and dNTP is edded, whereln the dNTP comprises a blocking molety (to facilitate removal of unextended primers) or e hapten to allow purification of extended primer, I.e. biotin, DNP, fluorescein, etc. Figure 14B depicts the same reaction with the use of an adapter sequence **90**; in this embodiment, the seme edapter sequence **90** may be used for each readout probe for an ellele. Figure 14C depicts the use of differant adepter sequences **90** for each readout probe; In this ambodiment, unreacted primers nead not be removed, although they may be. Figure 14D deplcts a solid phese reection, wherein the dNTP edded in the position adjacent to the reedout position **35** is labeled.

Figures 15A and 15B depict genotyping essays based on the novel combination of invasive cleavage end ligation reections. Figure 15A is e solution reaction, with the signalling probe **60** comprising e detection sequence **65** with e detectable label **45**. After hybridization with tha target sequence **25** and cleevege, the free detection sequence can bind to an arrey (deplcted hereln es a bead erray, although any nucleic ecid array can be used), using a capture probe **20** and e templete target sequence **26** for the ligation reaction. In the absence of ligation, the signalling probe is washed away. Figure 15B

depicts a solid phase essey. In this embodiment, the 5' end of the signelling probe is etteched to the erray (agein, depicted herein es e beed erray, elthough eny nucleic ecid errey cen be used), end e blocking molety is used et the 3' end. After cleavage, e free 3' end is generated, that cen then be used for ligation using e template target **26**. As will be appreciated by those in the ert, the orientation of this mey be switched, such that the 3' end of the signalling probe **60** is attached, end e free 5' end is genereted for the ligetion reaction.

Figures 16A and 16B depict genotyping essays besed on the novel combinetion of invasive cleavage and extension reactions. Figure 16A depicts en initial solution based assey, using a signalling probe with a blocked 3' end. After cleavage, the detection sequence can be edded to en erray end e templete target added, followed by extension to add e detectable lebel. Altemetively, the extension can elso heppen in solution, using e template target **26**, followed by additon of the extended probe to the erray. Figure 16B depicts the solid phase reaction; as above, either the 3' or the 5' end can be etteched. By using a blocking moiety **47**, only the newly cleaved ends may be extended.

Figures 17A, 17B and 17C depict three configurations of the combination of ligation end extension ("Genetic Bit" anelysis) for genotyping. Figure 17A depicts e reaction wherein the capture probe **20** and the extension probe serve es two ligetion probes, and hybridize adjacently to the terget sequence, such thet an edditionel ligation step mey be done. A labeled nucleotide is added et the reedout position. Figure 17B depicts a preferred embodiment, wherein the ligetion probes (one of which is the cepture probe **20**) ere separated by the detection position **30**. The addition of e labeled dNTP, extension enzyme and ligase thus serve to detect the readout position. Figure 17C depicts the solution phase assey. As will be eppreciated by those in the art, an extra level of specificity is added if the capture probe **20** spens the ligated probe **80**.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the detection end quantification of a veriety of nucleic ecid reactions, particularly using microsphere arrays. In particular, the invention relates to the detection of amplification, genotyping, end sequencing reactions. In eddition, the invention can be utilized with adapter sequences to create universal errays.

Accordingly, the present invention provides compositions and methods for detecting and/or quantifying the products of nucleic acid reections, such es terget nucleic ecid sequences, in a sample. As will be appreciated by those in the ert, the semple solution may comprise eny number of things, including, but not limited to, bodily fluids (including, but not limited to, blood, urine, serum, lymph, saliva, enal and vaginal secretions, perspiration and semen, of virtually eny orgenism, with mammalian samples being preferred end human samples being particularly preferred); environmental samples (including, but not limited to, air, egricultural, water and soil samples); biological warfare egent samples; research

samples; purified samples, such as purified genomic DNA, RNA, proteins, etc.; raw samples (bacteria, virus, genomic DNA, etc.; As will be eppreciated by those in the ert, virtually eny experimental manipulation may heva been done on the semple.

The present invention provides compositions end methods for detecting the presence or absence of target nucleic ecid sequences in a sample. By "nucleic acid" or "oligonucleotide" or grammaticel equivalents herein means at leest two nucleotides covalently linked together. A nucleic ecid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic ecid analogs ere included that may have eltemete backbones, comprising, for example, phosphoramide (Beeucege et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Ras. 14:3487 (1986); Sewal et al, Cham. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et el., Nucleic Acids Res. 19:1437 (1991); and U.S. Patent No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphophoroemidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier at al., Cham. Int. Ed. Engl. 31:1008 (1992); Nlelsen, Nature, 365:566 (1993); Cerisson et al., Neture 380:207 (1996), ell of which are incorporated by reference). Other enalog nucleic ecids include those with positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Patent Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 erid 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. Englisn 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrete Modifications In Antisense Research", Ed. Y.S. Sanghui and P. Den Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrehedron Lett. 37:743 (1996)) end non-ribose backbones, including those described in U.S. Petent Nos. 5,235,033 end 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modificetions in Antisense Research", Ed. Y.S. Senghui end P. Den Cook. Nucleic acids containing one or more carbocyclic sugers ere elso included within the definition of nucleic acids (see Jenkins et el., Chem. Soc. Rev. (1995) pp169-176). Sevarel nucleic ecid enelogs are described in Rewis, C & E News June 2, 1997 page 35. All of these references ere hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of labels, or to increase the stability end helf-life of such molecules in physiological environments.

As will be eppreciated by those in the art, all of these nucleic ecid enalogs may find use in the prasent invention. In eddition, mixtures of naturally occurring nucleic acids and enelogs can be made. Alternatively, mixtures of different nucleic acid analogs, end mixtures of naturally occuring nucleic acids and analogs may be made.

Particularly preferred ere peptide nucleic ecids (PNA) which includes peptide nucleic acid enalogs. These backbones ere substantially non-ionic under neutral conditions, in contrest to the highly charged phosphodiester backbone of neturally occurring nucleic acids. This results in two advantages. First, the PNA backbona exhibits improved hybridizetion kinetics. PNAs have larger changes in the melting temperature (Tm) for mismetched versus perfectly metched basepairs. DNA end RNA typically exhibit a 2-4°C drop in Tm for en internal mismatch. With tha non-ionic PNA beckbone, the drop is closer to 7-9°C. This allows for better detection of mismetches. Similarly, due to their non-ionic nature, hybridizetion of the bases attached to these backbones is relatively insensitive to salt concentration.

The nucleic acids mey be single strended or double stranded, es specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may ba DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains eny combinetion of deoxyribo- end ribo-nucleotides, end eny combination of beses, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc. A preferred embodiment utilizes isocytosine end isoguanine in nucleic ecids designed to be complementary to other probes, rether then target sequences, es this reduces non-specific hybridization, as is generally described in U.S. Patent No. 5,681,702. As used herein, the term "nucleoside" includes nucleotides es well as nucleoside end nucleotide enalogs, and modifiad nucleosides such as amino modified nucleosides. In eddition, "nuclaoside" includes non-naturally occuring analog structures. Thus for example the individual units of a peptide nucleic acid, each containing e base, are referred to herein as a nucleoside.

The compositions and methods of the invention ere directed to the detection of target sequences. The term "target sequence" or "target nucleic acid" or grammatical equivalents herein maans a nucleic acid sequence on a single strend of nucleic ecid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. As is outlined herein, the target sequence mey be a target sequence from a sample, or a secondery target such as a product of a reaction such es e detection sequence from an invasive cleavege reaction, a ligated probe from an OLA reaction, en extended probe from a PCR or SBE reaction, etc. Thus, for example, a target sequence from e sample is amplified to produce a secondary target that is detected; eltemetively, en emplification stap is done using a signal probe that is amplified, egeln producing a secondary target that is detacted. The target sequence mey be any length, with the understanding that longer sequences ere more specific. As will be eppreciated by those in the art, the complementary target sequenca mey take meny forms. For example, it may be contained within a larger nucleic ecid sequence, i.e. all or part of e gene or mRNA, e restriction fragment of a plasmid or genomic DNA, among others. As is outlined mora fully below, probes ere made to hybridize to target sequences to determine the presence, ebsence or quantity of a target sequence in a sample. Generally speeking, this term will be understood by those skilled in the ert. The target sequence may also be comprised of different target domains; for exampla, in "sendwich" type essays as outlined below, e first terget domain of the sample target sequence may hybridize to a cepture probe or e

portion of capture extender proba, e second terget domeln may hybridlza to a portion of en amplifier probe, e label probe, or a different capture or capture extender probe, etc. In addition, tha target domelns may be adjacent (l.a. contiguous) or separated. For example, when OLA techniques are used, a first primer may hybridiza to a first targat domain and a sacond primar may hybridiza to a second terget domain; either the domeins are adjecent, or they mey be separated by one or more nucleotides, coupled with the use of a polymerase end dNTPs, as is more fully outlined below. The terms "first" and "second" are not meant to confer an orientation of the sequances with respect to the 5'-3' orientation of the target sequence. For example, essuming a 5'-3' orientation of the complementary target sequence, tha first target domain mey be located either 5' to the second domain, or 3' to the second domain. In eddition, as will be appreciated by those in the art, tha probes on the surface of the erray (e.g. attached to the microspheres) may be attached in either orientation, either such that thay have a free 3' end or a frea 5' end; in some embodiments, the probes can be ettached at ona ora more internal positions, or at both ends.

If required, the target sequence is prepared using known techniques. For exampla, tha sample may be treated to lyse the cells, using known lysis buffers, sonication, electroporation, etc., with purification and amplification as outlined below occurring as needed, as will be eppreciated by those in the art. In addition, the reactions outlined hereln mey be eccomplished in a variety of ways, as will be appreciated by those in the art. Components of the reaction may be added simultaneously, or saquentially, in eny order, with preferred embodiments outlined below. In addition, tha reaction may include a variety of other reagents which may be included in the assays. Thesa include reagents like salts, buffers, neutral proteins, e.g. albumin, detergents, etc., which mey be used to facilitate optimal hybridization end detection, and/or reduce non-specific or background interactions. Also reagents that otherwisa improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial egents, etc., may be used, depending on the sampla preparation methods and purity of the target.

In addition, in most embodiments, double strended target nucleic ecids are denatured to render them single stranded so as to permit hybridization of the primers and other probes of the invention. A preferred embodiment utilizes a thermel step, generally by raising tha temperature of the reaction to ebout 95°C, although pH chenges end other techniques mey also be used.

As outlined herein, the Invention provides a number of differant primars and probes. By "primer nucleic acid" herein is meent a probe nucleic acid thet will hybridize to some portion, i.e. e domeln, of the target saquance. Probes of the present Invantion ere designed to be complementary to a target saquence (either the target sequance of the sample or to other probe sequences, as is described below), such that hybridizetion of the terget sequence and the probes of the prasent Invention occurs. As outlined below, this complemantarity need not be perfect; there may be any number of base pair mismatches which will interfere with hybridizetion between the target sequence end the single

stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under normal reaction conditions.

A variety of hybridization conditions may be used in the present invention, including high, moderate and low stringency conditions; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al, hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of helix destabilizing agents such as formamide. The hybridization conditions may also vary when a non-ionic backbone, i.e. PNA is used, as is known in the art. In addition, cross-linking agents may be added after target binding to cross-link, i.e. covalently attach, the two strands of the hybridization complex.

Thus, the assays are generally run under stringency conditions which allows formation of the hybridization complex only in the presence of target. Stringency can be controlled by altering a step parameter that is a thermodynamic variable, including, but not limited to, temperature, formamide concentration, salt concentration, chaotropic salt concentration, pH, organic solvent concentration, etc.

These parameters may also be used to control non-specific binding, as is generally outlined in U.S. Patent No. 5,681,697. Thus it may be desirable to perform certain steps at higher stringency conditions to reduce non-specific binding.

The size of the primer nucleic acid may vary, as will be appreciated by those in the art, in general varying from 5 to 500 nucleotides in length, with primers of between 10 and 100 being preferred, between 15 and 50 being particularly preferred, and from 10 to 35 being especially preferred,

depanding on the use and emplification technlque.

In addition, tha different amplification techniques may have further requirements of the primers, es is more fully described below.

In addition, es outlined herein, a variety of labeling techniques can be done.

Lebeling techniques

In general, either direct or indirect detection of tha target products can be done. "Direct" detection as used in this context, es for the other reactions outlined hereln, requires the incorporetion of e lebel, in this case e detectebla label, praferably an optical label such as a fluorophore, into the target sequence, with detection proceeding es outlined below. In this embodiment, the label(s) may be incorporated in e variety of ways: (1) the primers comprisa tha label(s), for example attached to the base, a ribosa, e phosphate, or to enalogous structures in e nucleic acid analog; (2) modified nucleosides are used that are modified et aither the base or the ribose (or to enelogous structures in a nucleic acid analog) with the label(s); these label-modified nucleosides are then converted to the triphosphate form and are incorporated into a nawly synthesized strand by a polymerasa; (3) modified nucleotides ere used that comprise a functional group thet cen be used to add a detectable label; (4) modified primers ara used that comprise a functional group that can be used to add a detectable label or (5) a label probe that is directly lebeled and hybridizes to a portion of the target sequence can be used. Any of these methods result in a newly synthesized strand or raaction product that comprises labels, thet can be directly detected as outlined below.

Thus, the modified strands comprise a detection label. By "detection label" or "detecteble label" herein is meant e moiety that ellows detection. This may be a primary label or a secondery label. Accordingly, detection labels may be primary lebels (i.a. directly detectable) or secondary labals (indirectly detectable).

In a preferred embodiment, the detection label is a primary label. A primary label is one that can be directly detected, such as a fluorophore. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heevy isotopes; b) magnetic, electrical, thermal labels; end c) colored or luminescent dyes. Labels can also include enzymes (horseradish peroxidase, etc.) and magnetic particles. Preferred labels include chromophores or phosphors but are preferably fluorescent dyes. Suitable dyes for use in the invention include, but ara not limited to, fluorescent lanthanide complexes, including those of Europium end Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumerins, quantum dots (elso referred to as "nanocrystals": see U.S.S.N. 09/315,584, hereby incorporated by reference), pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, Cy dyes (Cy3, Cy5, etc.), alexa dyes, phycoarythin, bodipy, end others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Heugland, hereby

expressly incorporated by reference.

In a prefarred embodimant, a sacondary detactable labal is used. A sacondary labal is ona that is indirectly detected; for example, a secondary label can bind or react with e primary label for detection, can act on an additional product to generate e primary lebel (e.g. enzymes), or may allow the seperetion of the compound comprising tha secondary label from unlabalad matarials, atc. Sacondary labels find particular usa in systems requiring separation of labalad and unlebaled probes, such as SBE, OLA, invasiva cleavaga reactions, etc; in eddition, these techniques may be usad with many of tha othar techniques dascribad harain. Sacondary labels include, but ara not limitad to, ona of a binding partner pair; chamically modifiabla moietias; nucleasa inhibitors, anzymes such as horsaradish peroxidasa, alkalina phosphatases, lucifierases, etc.

In a preferred embodimant, tha sacondary labal is a binding partner pair. For example, the label may be a hapten or antigan, which will bind its binding partner. In a preferred embodiment, the binding partner can be attached to a solid support to ellow separetion of extended and non-extanded primers. For example, suitable binding partner pairs include, but ere not limited to: antigens (such as proteins (including peptidas)) and antibodies (including fragmants tharaof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin; enzymas and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partnars. Nuclaic acid - nucleic acid binding proteins pairs ere also useful. In general, tha smaller of the pair is attached to the NTP for incorporation into the primer. Praferred binding partner pairs include, but ere not limited to, biotin (or imino-biotin) and streptavidin, digeoxinin and Abs, and Prolinx™ reagants (see www.prolinxinc.com/ie4/home.hmtl).

In a preferred embodiment, the binding partnar pair comprises biotin or imino-biotin and streptavidin. Imino-biotin is particularly prafarrad as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh danaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95°C).

In a preferred embodiment, the binding partnar pair comprises a primary detection label (for example, attached to tha NTP and therefora to the extended primar) and an antibody that will specifically bind to the primary detection label. By "spacifically bind" herein is meant that the partnars bind with spacificity sufficiant to differantiate between the pair and other componants or contaminants of the system. The binding should be sufficiant to ramain bound under the conditions of the assay, including wash staps to remova non-spacific binding. In soma embodiments, the dissociation constants of the pair will be less than about 10^{-4}-10^{-6} M^{-1}, with less than about 10^{-5} to 10^{-9} M^{-1} baing preferrad and less than about 10^{-7} - 10^{-9} M^{-1} baing particularly prefarrad.

In e praferrad embodiment, the sacondary label is a chemically modifiabla molaty. In this embodiment, labels comprising raactiva functional groups ere incorporated into the nucleic acid. Tha

functional group can then be subsequently labeled with a primary label. Suitable functional groups include, but ere not limited to, amino groups, carboxy groups, meleimide groups, oxo groups end thiol groups, with amino groups and thiol groups being particularly preferrad. For example, primary labels containing amino groups can be attached to secondary labels comprising amino groups, for exemple using linkers as are known in the art; for exemple, homo-or hetero-bifunctionel linkers es ere well known (see 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated herein by reference).

For removal of unextended primers, it is preferred that the other helf of the binding pelr is ettached to a solid support. In this embodiment, the solid support may be eny as described herein for substrates end microspheras, end the form is preferably microspheres es well; for example, a preferred embodiment utilizes magnetic beads that can be eesily introduced to the eempla end easily removad, although any affinity chromatography formats may be used as well. Standard methods are used to attach the binding partner to the solid support, end can include direct or indirect attachment methods. For exemple, biotin labeled antibodies to fluorophores cen be attached to streptevidin coeted magnetic beads.

Thus, in this embodiment, the extended primers comprise e binding partner that is contacted with its binding partner under conditions wherein the extended or reacted primers are separated from the unextended or unreacted primers. Thesa modified primers can then be added to the arrey comprising cepture probes as described herein.

Removal of unextended primers

In a preferred embodiment, it is desirable to remove the unextanded or unreected primers from the essay mixture, and particularly from the array, es unextended primers will compete with the extended (labeled) primers in binding to cepture probes, thereby diminishing the signal. The concentration of the unextended primers relative to the extended primer may be reletively high, since e large excess of primer is usually required to generate efficient primer annealing. Accordingly, a number of different techniques mey be used to facilitate the removal of unextended primers. While the discussion below applies specifically to SBE, these techniques may be used in any of the methods described herein.

In a preferred embodiment, the NTPs (or, in the case of other methods, one or more of the probes) comprise e secondary detectable label that can be used to separate extended and non-extended primers. As outlined ebove, detection labels may be primary labels (i.a. directly detectable) or secondary labels (indirectly detectable). A secondary label is one that is indirectly datected; for example, e secondery lebel can bind or reect with a primary label for detection, or may ellow the separation of the compound comprising the secondary label from unlebeled materiels, etc. Secondary labels find particular use in systems requiring separation of labeled end unlabeled probes, such es SBE, OLA, invasive cleevage, etc. reactions; in addition, these techniques may be used with many of

the other techniques described herein. Secondary labels include, but are not limited to, one of e binding partner pelr; chemically modifiable moleties; nuclease inhibitors, etc.

In a prafarred embodiment, the secondary label is a binding partner pair as outlined above. In a preferred embodiment, the binding partner pair comprises biotin or Imino-biotin and streptavidin. Imino-biotin is particularly preferrad when the methods require the leter separation of the palr, as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamida at 95°C).

In addition, the use of streptevidin/biotin systems can be used to separate unreacted and reacted probes (for axample in SBE, invasive cleavage, etc.). For example, the eddition of streptavidin to e nucielc ecid greatly increases its siza, as well as changes its physical properties, to allow mora efficient separetion techniques. For example, the mixtures can be size frectionated by exclusion chromatography, affinity chromatography, filtration or differential precipitation. Alternatively, an 3' axonuclease may be added to a mixture of 3' labeled biotin/streptavidin; only the unreacted oligonucleotides will be degraded. Following exonucleasa traatment, the exonuclaase and the streptavidin can be degraded using a protease such es proteinase K. The surviving nucleic acids (i.e. those that were biotinylated) are then hybridized to the array.

In a preferred embodiment, the binding partner pair comprises a primary detection label (attached to the NTP and therefore to the extended primer) and en antibody that will specificelly bind to the primery detection label.

In this embodiment, it is preferred that the other half of the binding pair is attached to a solid support. In this embodiment, the solid support may be any as described herein for substrates and microspheres, and the form is preferably microspheres as well; for example, a preferred embodiment utilizes magnetic beads that can be easlly introduced to the sample end easily removed, although any affinity chromatography formats may be used as well. Standard methods are used to ettach the binding partner to the solid support, and can include direct or indirect attachment methods. For axample, biotin labeled antibodies to fluorophores can be attached to streptavidin coated magnetic beads.

Thus, in this embodiment, tha axtended primers comprise a binding member that is contacted with its binding partner under conditions wherein the extended primers are separated from the unextended primers. These extended primers can then be added to the erray comprising capture probes as described herein.

In a preferred embodiment, the secondery label is a chemically modifiable moiety. In this embodiment, lebels comprising reactive functional groups are incorporated into the nucleic acid.

In a preferred embodiment, the secondary label is e nuclease inhibitor. In this embodiment, the chain-terminating NTPs are chosen to render extended primers resistant to nucleeses, such as 3'-exonucleases. Addition of en exonuclease will digest the non-extended primers leeving only the extended primers to bind to the capture probes on the erray. This may also be done with OLA, wherein the ligated probe will be protected but the unprotected ligation probe will be digested.

In this embodiment, suitable 3'-exonucleases include, but ere not limited to, exo I, exo III, exo VII, etc.

The present invention provides e variety of amplification reactions that can be detected using the arrays of the invention.

AMPLIFICATION REACTIONS

In this embodiment, the invention provides compositions end methods for the detection (end optionally quantification) of products of nucleic acid amplification reactions, using bead arrays for detection of the amplification products. Suitable emplificetion methods include both target amplification and signal amplification end include, but are not limited to, polymerase chain reaction (PCR), ligation chain reaction (sometimes referred to as oligonucleotide ligase amplification OLA), cycling probe technology (CPT), strand displacement essay (SDA), transcription medieted amplification (TMA), nucleic ecid sequence based amplification (NASBA), rolling circle amplification (RCA), end invasive cleavage technology. All of these methods require a primer nucleic acid (including nucleic acid analogs) that is hybridized to a target sequence to form a hybridization complex, and an enzyme is added that in some way modifies the primer to form e modified primer. For example, PCR generally requires two primers, dNTPs and a DNA polymerase; LCR requires two primers that adjacently hybridize to the target sequence and a ligase; CPT requires one cleavable primer end a cleaving enzyme; invesive cleavage requires two primers end a cleavage enzyme; etc. Thus, in general, a target nucleic acid is added to a reaction mixture that comprises the necessary emplification components, and a modified primer is formed.

In general, the modified primer comprises a detectable lebel, such es a fluorescent label, which is either incorporated by the enzyme or present on the original primer. As required, the unreacted primers are removed, in e variety of ways, as will be appreciated by those in the art and outlined herein. The hybridization complex is then disessociated, and the modified primer is detected end optionally quentitated by e microsphere array. In some cases, the newly modified primer serves es a target sequence for a secondary reaction, which then produces a number of emplified strands, which can be detected as outlined herein.

Accordingly, the reaction starts with the addition of e primer nucleic acid to the target sequence which

forms a hybridizetion complex. Once the hybridization complex between the primer end the target sequence hes been formed, en enzyme, sometimes termed en "amplification enzyme", is used to modify the primer. As for ell the methods outlined herein, the enzymes may be edded at eny point during the essay, either prior to, during, or after the addition of the primers. The identity of the enzyme will depend on the emplification technique used, as is more fully outlined below. Simllerly, the modificetion will depend on the amplification technique, es outlined below.

Once the enzyme hes modified the primer to form a modified primer, the hybridizetion complex is disessociated. In one espect, dissociation is by modification of the assay conditions. In enother espect, the modified primer no longer hybridizes to the target nucleic ecid end dissociates. Either one or both of these espects cen be employed in signel end target emplification reactions es described below. Generally, the amplification steps are repeated for a period of time to allow a number of cycles, depending on the number of copies of the original target sequence and the sensitivity of detection, with cycles ranging from 1 to thousends, with from 10 to 100 cycles being preferred and from 20 to 50 cycles being especially preferred.

After a suitable time of amplification, unreacted primers are removed, in a variety of ways, es will be appreciated by those in the art and described below, and the hybridization complex is disassociated. In general, the modified primer comprises a detectable label, such es a fluorescent label, which is either incorporated by the enzyme or present on the original primer, and the modified primer is added to a microsphere array such is generelly described in U.S.S.N.s 09/189,543; 08/944,850; 09/033,462; 09/287,573; 09/151,877; 09/187,289 and 09/256,943; and PCT applications US98/09163 and US99/14387; US98/21193; US99/04473 and US98/05025, ell of which are hereby incorporated by reference. The microsphere errey comprises subpopulations of microspheres that comprise capture probes that will hybridize to the modified primers. Detection proceeds via detection of the label as an indication of the presence, absence or amount of the target sequence, as is more fully outlined below.

TARGET AMPLIFICATION

In e preferred embodiment, the amplification is target emplification. Target amplification involves the amplification (replication) of the target sequence to be detected, such thet the number of copies of the target sequence is increesed. Suitable target emplification techniques include, but ere not limited to, the polymerase chain reaction (PCR), strand displacement amplification (SDA), transcription mediated amplification (TMA) and nucleic acid sequence based emplificetion (NASBA).

POLYMERASE CHAIN REACTION AMPLIFICATION

In e preferred embodiment, the target amplification technique is PCR. The polymerase chain reaction (PCR) is widely used end described, end involves the use of primer extension combined with thermal cycling to amplify a target sequence; see U.S. Patent Nos. 4,683,195 and 4,683,202, and PCR Essential Data, J. W. Wiley & sons, Ed. C.R. Newton, 1995, all of which are incorporated by reference.

In eddition, there are a number of variations of PCR which also find use in the invention, including "quantitative competitive PCR" or "QC-PCR", "erbitrarily primed PCR" or "AP-PCR" , "immuno-PCR", "Alu-PCR", "PCR single strand conformetional polymorphism" or "PCR-SSCP", "reverse transcriptase PCR" or "RT-PCR", "biotin capture PCR", "vectorette PCR", "panhendle PCR", end "PCR select cDNA subtraction", "allele-specific PCR", emong others. In some embodiments, PCR is not preferred.

In general, PCR may be briefly described es follows. A double strended target nucleic ecid is denetured, generally by reising the temperature, end then cooled in the presence of an excess of e PCR primer, which then hybridizes to the first terget strand. A DNA polymerase then ects to extend the primer with dNTPs, resulting in the synthesis of a new strand forming e hybridizetion complex. The sample is then heated egein, to disassociate the hybridizetion complex, end the process is repeated. By using e second PCR primer for the compiementary target strand, rapid and exponentiel amplification occurs. Thus PCR steps ere denaturation, annealing end extension. The particulars of PCR ere well known, and include the use of e thermostable polymerase such as Taq I polymerase and thermal cycling.

Accordingly, the PCR reaction requires at least one PCR primer, e polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

In general, as is more fully outlined below, the capture probes on the beads of the array are designed to be substantially compiementary to the extended part of the primer; that is, unextended primers will not bind to the capture probes. Alternatively, as further described below, unreacted probes may be removed prior to addition to the array.

STRAND DISPLACEMENT AMPLIFICATION (SDA)

In a preferred embodiment, the target amplification technique is SDA. Strend displacement amplificetion (SDA) is generally described in Walker et al., in Molecular Methods for Virus Detection, Academic Press, Inc., 1995, and U.S. Patent Nos. 5,455,166 end 5,130,238, ell of which are hereby expressly incorporated by reference in their entirety.

In general, SDA may be described as follows. A single stranded target nucleic ecid, usually a DNA target sequence, is contected with an SDA primer. An "SDA primer" generally has a length of 25-100 nucleotides, with SDA primers of epproximately 35 nucleotides being preferred. An SDA primer is substantially complementary to a region at the 3' end of the target sequence, and the primer has a sequence at its 5' end (outside of the region that is complementary to the target) thet is a recognition sequence for a restriction endonucleese, sometimes referred to herein es e "nicking enzyme" or a "nicking endonuclease", as outlined below. The SDA primer then hybridizes to the target sequence. The SDA reaction mixture also contains e polymerase (an "SDA polymerase", es outlined below) end

e mixture of all four deoxynucleoside-triphosphates (also called deoxynucleotides or dNTPs, i.e. dATP, dTTP, dCTP and dGTP), at least one species of which is a substituted or modified dNTP; thus, the SDA primer is modified, i.e. extended, to form a modified primer, sometimes referred to herein as a "newly synthesized strand". The substituted dNTP is modified such that it will inhibit cleavage in the strand containing the substituted dNTP but will not inhibit cleavage on the other strand. Examples of suitable substituted dNTPs include, but are not limited, 2'deoxyadenosine 5'-O-(1-thiotriphosphate), 5-methyldeoxycytidine 5'-triphosphate, 2'-deoxyuridine 5'-triphosphate, adn 7-deaza-2'-deoxyguanosine 5'-triphosphate. In addition, the substitution of the dNTP may occur after incorporation into a newly synthesized strand; for example, a methylase may be used to add methyl groups to the synthesized strand. In addition, if all the nucleotides are substituted, the polymerase may have 5'→3' exonuclease activity. However, if less than all the nucleotides are substituted, the polymerase preferably lacks 5'→3' exonuclease activity.

As will be appreciated by those in the art, the recognition site/endonuclease pair can be any of a wide variety of known combinations. The endonuclease is chosen to cleave a strand either at the recognition site, or either 3' or 5' to it, without cleaving the complementary sequence, either because the enzyme only cleaves one strand or because of the incorporation of the substituted nucleotides. Suitable recognition site/endonuclease pairs are well known in the art; suitable endonucleases include, but are not limited to, HincII, HindII, AvaI, Fnu4HI, TthIII, NciI, BstXI, BamHI, etc. A chart depicting suitable enzymes, and their corresponding recognition sites and the modified dNTP to use is found in U.S. Patent No. 5,455,166, hereby expressly incorporated by reference.

Once nicked, a polymerase (an "SDA polymerase") is used to extend the newly nicked strand, 5'→3', thereby creating another newly synthesized strand. The polymerase chosen should be able to intiate 5'→3' polymerization at a nick site, should also displace the polymerized strand downstream from the nick, and should lack 5'→3' exonuclease activity (this may be additionally accomplished by the addition of a blocking agent). Thus, suitable polymerases in SDA include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase.

Accordingly, the SDA reaction requires, in no particular order, an SDA primer, an SDA polymerase, a nicking endonuclease, and dNTPs, at least one species of which is modified. Again, as outlined above for PCR, preferred embodiments utilize capture probes complementary to the newly synthesized portion of the primer, rather than the primer region, to allow unextended primers to be removed.

In general, SDA does not require thermocycling. The temperature of the reaction is generally set to be high enough to prevent non-specific hybridization but low enough to allow specific hybridization; this is generally from about 37°C to about 42°C, depending on the enzymes.

In a preferred embodiment, as for most of the amplification techniques described herein, a second amplification reaction can be done using the complamentary target sequence, resulting in a substantiel increase in emplification during e set period of time. That is, e second primer nucleic ecid is hybridized to a second target sequence, that is substantially complementary to the first target sequence, to form a second hybridization complex. The addition of tha enzyme, followed by disessociation of the sacond hybridizetion complex, results in the generation of a number of newly synthesized second strands.

NUCLEIC ACID SEQUENCE BASED AMPLIFICATION (NASBA) AND TRANSCRIPTION MEDIATED AMPLIFICATION (TMA)

In a preferred embodiment, the target amplification technique is nucleic acid sequence based amplification (NASBA). NASBA is generally described in U.S. Patent No. 5,409,818; Sooknanan et al., Nucleic Acid Sequence-Based Amplificetion, Ch. 12 (pp. 261-285) of Molecular Methods for Virus Detection, Academic Press, 1995; and "Profiting from Gene-based Diagnostics", CTB International Publishing Inc., N.J., 1996, all of which are incorporated by reference. NASBA is very similar to both TMA and QBR. Transcription mediated amplification (TMA) is generally described in U.S. Patent Nos. 5,399,491, 5,888,779, 5,705,365, 5,710,029, all of which are incorporated by reference. Tha main difference between NASBA and TMA is that NASBA utilizes the eddition of RNAse H to effect RNA degradation, end TMA relies on inherent RNAse H activity of tha reverse trenscriptase.

In general, these techniques mey be described es follows. A single stranded target nucleic acid, usually an RNA target sequence (sometimes referred to herein as "the first target sequence" or "the first tamplate"), is contacted with a first primer, generally referred to herein as a "NASBA primer" (although "TMA primer" is also suitable). Starting with a DNA target sequence is described below. These primers generally have a length of 25-100 nucleotides, with NASBA primers of approximately 50-75 nucleotides being preferred. The first primer is preferably a DNA primer that has at its 3' end a sequence that is substantially complementary to the 3' end of the first template. The first primer also has an RNA polymerase promoter at its 5' end (or its complement (antisense), depending on the configuration of the system). The first primer is then hybridized to the first templete to form e first hybridization complex. The reaction mixture also includes e reverse transcriptase enzyme (en "NASBA reverse trenscriptase") end e mixture of the four dNTPs, such that the first NASBA primer is modified, i.e. extended, to form a modified first primer, comprising a hybridization complex of RNA (the first template) and DNA (the newly synthesized strand).

By "reverse transcriptase" or "RNA-directed DNA polymerese" herein is meant en enzyme capable of synthesizing DNA from a DNA primer end an RNA template. Suitable RNA-directed DNA polymerases include, but are not limited to, evian myloblastosis virus reverse transcriptase ("AMV RT") and the Moloney murine leukemia virus RT. When the amplification reaction is TMA, the reverse transcriptase enzyme further comprises a RNA degrading activity as outlined below.

In addition to the components listed ebove, the NASBA reaction also includes an RNA degreding enzyme, elso sometimes referred to herein as e ribonuclease, that will hydrolyze RNA of en RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, but are not limited to, RNase H from *E. coli* and celf thymus.

The ribonuclease activity degrades the first RNA template in the hybridization complex, resulting in e disassociation of the hybridizetion complex leaving a first single stranded newly synthesized DNA strand, sometimes referred to herein as "the second template".

In addition, the NASBA reection also includes a second NASBA primer, generelly comprising DNA (although as for ell the probes herein, including primers, nucleic acid analogs may also be used). This second NASBA primer has a sequence at its 3' end that is substantially complementary to the 3' end of the second template, and elso contains an entisense sequence for a functional promoter and the entisense sequence of a trenscription initiation site. Thus, this primer sequence, when used as a template for synthesis of the third DNA template, contains sufficient information to allow specific and efficient binding of an RNA polymerase and initiation of trenscription at the desired site. Preferred embodiments utilizes the antisense promoter and transcription initiation site ere that of the T7 RNA polymerase, although other RNA polymerase promoters and initiation sites can be used as well, as outlined below.

The second primer hybridizes to the second template, and a DNA polymerase, also termed a "DNA-directed DNA polymerase", elso present in the reaction, synthesizes a third template (a second newly synthesized DNA strand), resulting in second hybridization complex comprising two newly synthesized DNA strands.

Finally, the inclusion of an RNA polymerese and the required four ribonucleoside triphosphates (ribonucleotides or NTPs) results in the synthesis of an RNA strand (a third newly synthesized strand that is essentielly the same es the first template). The RNA polymerese, sometimes referred to herein es a "DNA-directed RNA polymerase", recognizes the promoter and specifically initiates RNA synthesis at the initiation site. In addition, the RNA polymerase preferebly synthesizes several copies of RNA per DNA duplex. Preferred RNA polymereses include, but are not limited to, T7 RNA polymerase, end other bacteriophage RNA polymerases including those of phege T3, phage ϕII, Salmonella phage sp6, or Pseudomonese phage gh-1.

In some embodiments, TMA end NASBA ere used with sterting DNA target sequences. In this embodiment, it is necessary to utilize the first primer comprising the RNA polymerese promoter and a DNA polymerase enzyme to generete a double stranded DNA hybrid with the newly synthesized strand comprising the promoter sequence. The hybrid is then denatured end the second primer edded.

Accordingly, the NASBA reaction requires, in no particular order, e first NASBA primer, a second NASBA primer comprising an antisense sequence of an RNA polymarase promoter, an RNA polymerase thet recognizes the promoter, e revarse transcriptasa, a DNA polymerase, en RNA degrading enzyme, NTPs and dNTPs, in eddition to the detection components outlined below.

These components result in a single starting RNA template generating e single DNA duplex; however, since this DNA duplex results in the creation of multiple RNA strands, which can then be used to initiate the reaction egain, amplificetion proceeds rapidly.

Accordingly, the TMA reaction requires, in no particular order, a first TMA primer, a second TMA primer comprisIng en antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase with RNA degrading activity, a DNA polymerase, NTPs end dNTPs, in eddition to the detection components outlined below.

These components result in a single starting RNA template genereting e single DNA duplex; however, since this DNA duplex results in the creation of multipla RNA strands, which can then be used to initiate the reaction agein, amplification proceeds rapidly.

As outlined herein, the detection of the newly synthesized strands can proceed In several ways. Direct detection can be done when the newly synthesized strands comprisa detectable labels, either by incorporation into the primers or by incorporation of modified labelled nucleotides into the growing strand. Alternatively, as is more fully outlined below, indirect detection of unlabelled strands (which now serve as "targets" in the detection mode) can occur using a variety of sandwich assay configurations. As will be eppreciated by those in the art, eny of the newly synthasized strends can serve as the "target" for form an essay complex on a surface with a capture probe. In NASBA and TMA, it is preferable to utilize the newly formed RNA strands as the target, as this is where significant amplification occurs.

In this way, a number of secondary target molecules are made. As is more fully outlined below, these reactions (that is, the products of these reactions) can be detected in a number of ways.

SIGNAL AMPLIFICATION TECHNIQUES

In a preferred embodiment, the amplification technique is signal amplification. Signal emplificetion involves the use of limited number of target molecules as templates to either generate multiple signalling probas or ellow the use of multiple signalling probes. Signal amplification strategies include LCR, CPT, QβR, invasiva cleavaga technology, and the use of amplification probes in sandwich assays.

SINGLE BASE EXTENSION (SBE)

In a preferred embodiment, single base extension (SBE; sometimes referred to as "minisequencing") is used for amplification. It should also be noted that SBE finds use in genotyping, es is described below. Briefly, SBE is a technique that utilizes en extension primer thet hybridizes to the target nucleic acid. A polymerase (generally a DNA polymerase) is used to extend the 3' end of the primer with e nucleotide analog labeled a detection label es described hereln. Besed on the fidelity of the anzyme, a nucleotide is only incorporated into the extension primer if it is complementary to the adjacent basa in the target strand. Generelly, the nucleotide is derivatized such that no further extensions can occur, so only a singla nucleotide is added. However, for emplificetion raactions, this may not be necessary. Once the labeled nucleotide is added, detection of the lebel proceeds as outlined hereln. See generally Sylvanen et al., Genomics 8:684-692 (1990); U.S. Patent Nos. 5,846,710 and 5,888,819; Pastinen et al., Genomics Res. 7(6):606-614 (1997); all of which are expressly incorporated herein by reference.

The reaction is initiated by introducing the assay complex comprising the target sequence (i.e. the array) to a solution comprising a first nucleotide, frequently an nucleotide analog. By "nucleotida enelog" in this context herein is meant a deoxynucleoside-triphosphate (also called deoxynucleotides or dNTPs, i.e. dATP, dTTP, dCTP and dGTP), that is further derivatized to be chain terminating. As will be appreciated by those in the art, any number of nucleotide analogs may be used, as long as a polymerase enzyme will still incorporate the nucleotide at the interrogation position. Preferred embodiments utilize dideoxy-triphosphate nucleotides (ddNTPs). Generally, a set of nucleotides comprising ddATP, ddCTP, ddGTP and ddTTP is used, at least one of which includes a label, end preferably ell four. For amplification rather than genotyping reactions, the labels may all be the same; elternatively, different labels may be used.

In a preferred embodiment, the nucleotide analogs comprise a detectable label, which cen be either a primary or secondary datecteble label. Preferred primary labels are those outlined above. Howaver, the enzymatic incorporation of nucleotides comprising fluorophores is poor under meny conditions; accordingly, preferred embodiments utilize secondary detectable labels. In addition, es outlined below, the use of secondary labels may also facilitate the removal of unextended probes.

In addition to a first nucleotide, the solution also comprises en axtension enzyme, generelly e DNA polymerase. Suitable DNA polymerases include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase. If the NTP is complementary to tha base of the detection position of the target sequence, which is adjacent to the extension primer, the extension enzyma will add it to the extension primer. Thus, the extension primer is modified, i.e. extended, to form e modified primer, sometimes referred to herein as a "newly synthesizad strand".

A limitation of this method is thet unless the target nucleic acid is in sufficient concentration, the

amount of unextended primer in the reection greatly exceeds the resultant extended-labeled primer. The excess of unextended primer competes with the detection of the lebeled primer in the esseys described herein. Accordingly, when SBE is used, preferred embodiments utilize methods for the removal of unextended primers es outlined herein.

One method to overcome this limitation is thermocycling minisequencing in which repeated cycles of ennealing, primer extension, end heat denaturation using a thermocycler end thermo-stable polymerase allows the emplification of the extension probe which results in the eccumuletion of extended primers. For exemple, if the original unextended primer to target nucleic acid concentration is 100:1 end 100 thermocycles and extensions ere performed, a majority of the primer will be extended.

As will be appreciated by those in the ert, the configuration of the SBE system can take on several forms. As for the LCR reaction described below, the reaction may be done in solution, and then the newly synthesized strands, with the base-specific detectable labels, can be detected. For example, they can be directly hybridized to capture probes that ere complementary to the extension primers, and the presence of the label is then detected.

Alternatively, the SBE reaction can occur on a surface. For example, a target nucleic ecid may be captured using e first capture probe that hybridizes to e first target domain of the target, end the reection can proceed at a second target domain. The extended labeled primers ere then bound to a second capture probe end detected.

Thus, the SBE reaction requires, in no particular order, an extension primer, a polymerase end dNTPs, at least one of which is labeled.

OLIGONUCLEOTIDE LIGATION AMPLIFICATION (OLA)

In e preferred embodiment, the signal amplification technique is OLA. OLA, which is referred to es the ligation chain reaction (LCR) when two-stranded substrates ere used, involves the ligation of two smaller probes into a single long probe, using the target sequence es the template. In LCR, the ligated probe product becomes the predominant template as the reaction progresses. The method can be run in two different ways; in a first embodiment, only one strand of e target sequence is used as a template for ligetion; alternatively, both strands may be used. See generally U.S. Petent Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; WO 97/31256; end WO 89/09835, and U.S.S.N.s 60/078,102 end 60/073,011, all of which ere incorporated by reference.

In e preferred embodiment, the single-stranded target sequence comprises a first target domein end a second target domain, which ere adjacent end contiguous. A first OLA primer end a second OLA

primer nucleic ecids are added, that are substantially complementary to their respective target domain and thus will hybridize to the target domains. These target domeins may be directly adjecent, i.e. contiguous, or separated by a number of nuclaotides. If thay are non-contiguous, nuclaotidas are edded elong with means to join nucleotides, such as a polymerase, that will edd the nucleotides to one of the primers. The two OLA primers are then covalently attached, for exemple using e ligese enzyme such es is known in the ert, to form a modified primer. This forms e first hybridization complex comprising the ligated probe end the target sequence. This hybridization complex is then denatured (disassociatad), and the procass is rapeated to ganerate a pool of ligated probes.

In a preferred embodiment, OLA is done for two strands of e double-stranded target sequence. The target sequence is denetured, and two sets of probes are edded: one set es outlined above for one strand of the target, and e saperete set (i.e. third and fourth primer probe nucleic acids) for the other strand of the target. In a preferred embodiment, the first and third probes will hybridize, and tha second end fourth probes will hybridize, such that emplification cen occur. That is, when the first and second probes have been attached, the ligated probe can now be used as a template, in addition to the second target sequence, for the ettachment of the third and fourth probes. Similarly, the ligeted third and fourth probes will serve as a template for the ettechment of the first and second probes, in addition to the first target strand. In this wey, an exponential, rather than just a linear, amplification can occur.

As will be appreciated by those in the ert, the ligation product can be detected in e variety of ways. In a preferred embodiment, the ligation reaction is run in solution. In this embodiment, only one of the primers carries e detectable label, e.g. the first ligetion probe, end the cepture probe on the bead is substantially complementary to the other probe, e.g. the second ligation probe. In this way, unextended labeld ligation primers will not interfere with the assay. That is, in a preferred embodiment, the ligation product is detected by solid-phase oligonucleotide probes. The solid-phase probes ere preferably complementery to at least e portion of the ligation product. In e preferred embodiment, the solid-phase probe is complementary to the 5' detection oligonucleotide portion of the ligation product. This substantially reduces or eliminates felse signel generated by the optically-labeled 3' primers. Preferably, datection is accomplished by removing the unligated 5' detection oligonucleotida from tha reaction before application to a capture probe. In one embodiment, the unligated 5' detection oligonucleotides are removed by digesting 3' non-protected oligonucleotides with a 3' exonuclease, such es, exonuclease I. The ligation products are protected from exo I digestion by including, for example, 4-phosphorothioete residues at their 3' terminus, thereby, rendering them resistant to exonucleasa digastion. The unligated detection oligonuclaotidas ara not protected and ara digested.

Alternatively, the target nucleic acid is immobilized on a solid-phase surfece. The ligation essay is performed and unligated oligonucleotides ere removed by washing under eppropriate stringency to remove unligated oligonucleotides. The ligated oligonucleotides ere eluted from the target nucleic acid

using denaturing conditions, such es, 0.1 N NeOH, and detected as described herein.

Again, as outlined above, the detection of the LCR reaction can also occur directly, in the case whera one or both of the primers comprises at least one detectable label, or indirectly, using sandwich essays, through the use of additional probes; that is, the ligated probes can serve es target sequences, and detection may utilize amplification probes, capture probes, capture extander probes, label probes, and label extender probes, etc.

ROLLING-CIRCLE AMPLIFICATION (RCA)

In e preferred embodiment the signal amplification technique is RCA. Rolling-circla amplificetion is generally described in Baner a*t al.* (1998) *Nuc. Acids Res.* 26:5073-5078; Berany, F. (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193; and Lizardl e*t al.* (1998) *Net. Genet.* 19:225-232, all of whlch are incorporated by reference In their entirety.

In general, RCA may be described in two ways. First, as is outlined in more detail below, a single probe is hybridized with a target nucleic acid. Each terminus of the probe hybridizes adjacently on the target nucleic acid and the OLA essay as described above occurs. When ligated, the probe is circularized whila hybridized to the target nucleic acid. Addition of a polymerese results in extension of the circular proba. However, since the probe has no terminus, the polymerase continues to extend the probe repeetedly. Thus results in amplification of the circular probe.

A second alternative approach involves OLA followed by RCA. In this embodiment, en immobilized primer is contacted with a target nucleic acid. Complementary sequences will hybridize with each other resulting In an Immobilized duplex. A second primer is contacted with the target nucleic acid. The second primer hybridizes to the target nucleic acid adjacent to the first primer. An OLA assay Is performed es described above. Ligation only occurs if the primer are complementary to the target nucleic acid. When e mismatch occurs, particularly at one of the nucleotides to be ligated, ligation will not occur. Following ligation of the oligonucleotides, the ligeted, Immobilized, oligonucleotide Is then hybridized with an RCA proba. This is a circular probe that is desIgned to specifically hybridize with the ligated oligonucleotida and will only hybridize with an oligonucleotide thet has undergone ligation. RCA is than performed es is outlined in more detail balow.

Accordingly, in an preferrad embodiment, e single oligonucleotide is used both for OLA and as the circular tamplate for RCA (referred to herein as a "padlock probe" or a "RCA probe"). That is, aach terminus of the oligonucleotide contains sequence complementary to the target nucleic ecid and functions as an OLA primar es described above. That Is, the first and of the RCA probe is substantially complementery to e first target domaln, end the second end of the RCA probe is substantially complementary to a second target domain, adjecent to the first domaln. Hybridization of the oligonucleotide to the terget nucleic ecid results in the formation of a hybridization complex.

Ligation of the "primers" (which are the discrete ends of a single oligonucleotide) results in the formetion of e modified hybridizetion complex containing e circuler probe i.e. en RCA templete complex. That is, the oligonucleotide is circularized while still hybridized with the target nucleic ecid. This serves es a circuler templete for RCA. Addition of a polymerese to the RCA template complex results in the formetion of en emplified product nucleic ecid. Following RCA, the emplified product nucleic ecid is detected (Figure 6). This can be accomplished in a variety of ways; for example, the polymerase may incorporate labelled nucleotides, or elternatively, a label probe is used that is substantially complementary to a portion of the RCA probe and comprises at least one label is used.

The polymerase cen be eny polymerase, but is preferably one lacking 3' exonuclease activity (3' exo⁻). Examples of suitable polymerase include but are not limited to exonuclease minus DNA Polymerase I large (Klenow) Fragment, Phi29 DNA polymerase, Taq DNA Polymerase and the like. In eddition, in some embodiments, a polymerese that will replicate single-stranded DNA (i.e. without a primer forming a double stranded section) can be used.

In a preferred embodiment, the RCA probe contains an adapter sequence es outlined herein, with edepter cepture probes on the erray, for example on a microsphere when microsphere arrays are being used. Alternatively, unique portions of the RCA probes, for example all or part of the sequence corresponding to the target sequence, cen be used to bind to e capture probe.

In a preferred embodiment, the padlock probe contains a restriction site. The restriction endonuclease site allows for cleavage of the long concatamers that ere typically the result of RCA into smaller individual units that hybridize either more efficiently or faster to surface bound capture probes. Thus, following RCA, the product nucleic ecid is contacted with the appropriate restriction endonuclease. This results in cleavage of the product nucleic acid into smaller fragments. The fregments are then hybridized with the capture probe that is immobilized resulting in e concentration of product fragments onto the microsphere. Again, as outlined herein, these fragments can be detected in one of two ways: either labelled nucleotides are incorporated during the replication step, or an additional label probe is added.

Thus, in a preferred embodiment, the padlock probe comprises a label sequence; i.e. a sequence that can be used to bind label probes and is substantially complementary to e lebel probe. In one embodiment, it is possible to use the same label sequence and label probe for ell padlock probes on en array; elternatively, eech padlock probe can have a different label sequence.

The padlock probe elso contains a priming site for priming the RCA reaction. That is, each padlock probe comprises a aequence to which e primer nucleic acid hybridizes forming a template for the polymerese. The primer can be found in eny portion of the circular probe. In a preferred embodiment, the primer is located at a discrete site in the probe. In this embodiment, the primer site in each distinct

padlock probe is identical, although this is not required. Advanteges of using primer sites with identical eequences includa the ability to use only e single primer oligonucleotide to prime the RCA essey with a plurelity of different hybridization complexes. That is, the pedlock probe hybridizes uniquely to the target nucleic acid to which it is designed. A singla primer hybridizes to all of the unique hybridization complexes forming e priming site for the polymerase. RCA then proceeds from en identical locus within eech unique padlock probe of the hybridization complexes.

In an altemative embodiment, the primer site can overlap, encompass, or reside within any of the above-described elements of the padlock probe. That is, the primer can be found, for example, overlapping or within the restriction site or tha identifier sequence. In this embodiment, it is necessery that the primer nucleic acid is designed to base peir with the chosen primer site.

Thus, the padlock probe of the invention contains at each terminus, sequences corresponding to OLA primers. The intervening sequence of the padlock probe contain in no particular order, en adapter sequence and a restriction endonuclease site. In addition, the padlock probe contains a RCA priming site.

Thus, in a preferred embodiment the OLA/RCA is performed in solution followed by restriction endonuclease cleevage of the RCA product. The cleaved product is then applied to an array comprising beads, each bead comprising a probe complementary to the adapter sequence located in the padlock probe. The amplified edapter sequance correlates with a particular target nucleic acid. Thus the incorporation of en endonuclease site allows the generation of short, eesily hybridizable sequences. Furthermore, the unique adapter sequence in each rolling circle pedlock probe sequence allows diverse sets of nucleic acid sequences to be enalyzed in parallel on an array, since each sequence is resolved on the basis of hybridization specificity.

In en altemative OLA/RCA method, one of the OLA primers is immobilized on the microsphere; the second primer is edded in solution. Both primers hybridize with the target nucleic acid forming a hybridization complex as described ebove for the OLA assay.

As described herein, the microsphere is distributed on an arrey. In a preferred embodiment, a plurality of microspheres each with a unique OLA primer is distributed on the array.

Following the OLA assey, and either before, after or concurrently with distribution of the beads on the arrey, e segment of circular DNA is hybridized to the bead-based ligated oligonuclaotide forming e modified hybridizetion complex. Addition of an appropriate polymerase (3' exo'), as is known in the ert, and corresponding reection buffer to the arrey leads to emplification of the circular DNA. Since there is no terminus to the circuler DNA, the polymerase continues to travel around the circular template generating extension product until it detaches from the template. Thus, a polymerese with high

processivity can create saveral hundred or thousand copies of the circular templata with all the copies linked in one contiguous strand.

Again, these copies are subsequently detected by one of two methods; either hybridizing a labeled oligo complementary to the circular target or via the incorporation of labeled nucleotides in the amplification reection. The lebel is detected using conventionel label detection methods es described heraln.

In one embodiment, when the circular DNA contains sequences complementary to the ligated oligonucleotida it is preferable to remove the target DNA prior to contacting the ligeted oligonucleotide with the circular DNA (See Fig 7). This is done by denaturing tha double-stranded DNA by methods known in the art. In an alternative embodiment, the double stranded DNA is not denatured prior to contacting the circuler DNA.

In an elternative embodiment, when the circular DNA contains sequences complementary to the target nucleic ecid, it is preferable that the circular DNA is complementary at a site distinct from the site bound to the ligated oligonucleotide. In this embodiment it is preferred that the duplex between the ligated oligonucleotide and target nucleic acid is not denatured or disrupted prior to the addition of the circular DNA so that the target DNA remains immobilized to the bead.

Hybridization and washing conditions ere well known in the ert; various degrees of stringency can be used. In some embodiments it is not necessary to use stringent hybridization or washing conditions es only microspheres containing the ligated probes will effectively hybridize with tha circular DNA; microspheres bound to DNA that did not undergo ligation (those without the appropriate target nucleic ecid) will not hybridize as strongly with the circular DNA as those primers that ware ligated. Thus, hybridization and/or washing conditions ere used that discriminate between binding of the circular DNA to the ligated primer end the unligated primer.

Alternatively, when the circular probe is designed to hybridize to the target nucleic acid at a site distinct from the site bound to the ligated oligonucleotide, hybridization and washing conditions are used to remove or dissociate the target nucleic acid from unligated oligonucleotides while target nucleic ecid hybridizing with tha ligated oligonucleotides will remain bound to the beads. In this embodiment, the circuler probe only hybridizes to the target nucleic acid when the target nucleic ecid is hybridized with e ligated oligonucleotide that is immobilized on e bead.

As is well known in the art, an appropriate polymerase (3' exo$^-$) is added to the arrey. The polymerase extends the sequence of a single-stranded DNA using double-strended DNA es a primer site. In one embodiment, the circular DNA that has hybridized with the appropriate OLA reection product serves as the primer for the polymerase. In the presence of en appropriete reection buffer as is known in the ert,

the polymerasa will extend tha sequence of the primer using the single-stranded circuler DNA es a template. As there is no terminus of the circuler DNA, the polymerase will continue to extend the sequence of the circular DNA. In an alternative embodiment, the RCA probe comprises e discrete primer site located within the circular probe. Hybridization of primer nucleic acids to this primer site forms the polymerese templata allowing RCA to proceed.

In e preferred embodiment, the polymerase creates more than 100 copies of the circular DNA. In more preferred embodiments the polymerase creates more than 1000 copies of tha circular DNA; while in a most preferrad embodiment the polymerase creates more than 10,000 copies or more then 50,000 copies of the template.

The amplified circular DNA sequence is then detected by methods known in the art end as described herein. Detection is eccomplished by hybridizing with a labeled probe. The probe is labeled directly or indirectly. Alternatively, labeled nucleotides are incorporated into the emplified circular DNA product. The nucleotides can be labeled directly, or indirectly as is further described herein.

The RCA as described herein finds use in allowing highly specific and highly sensitive detection of nucleic acid target sequences. In particular, the mathod finds use in improving the multiplexing ebility of DNA arreys and eliminating costly semple or target preparation. As an example, a substantial sevings in cost can be realized by directly analyzing genomic DNA on an array, rather than employing an intermediate PCR amplification step. The method finds use in examining genomic DNA end other samples including mRNA.

In addition the RCA finds use in allowing rolling circle amplification products to be easily detected by hybridization to probes in a solid-phese format (a.g. en array of beeds). An additional advantage of the RCA is that it provides the capability of multiplex analysis so that large numbers of sequences can be analyzed in parallel. By combining the sensitivity of RCA and parallel detection on arrays, meny sequences cen be analyzed directly from genomic DNA.

CHEMICAL LIGATION TECHNIQUES

A variation of LCR utilizes a "chemical ligation" of sorts, as is generelly outlined in U.S. Patent Nos. 5,616,464 and 5,767,259, both of which are hereby expressly incorporated by reference in their entirety. In this embodiment, similar to enzymetic ligation, a pair of primers are utilized, wherein the first primer is substantially complementary to a first domein of the target end the second primer is substantially complementary to an edjecent second domain of the target (although, as for enzymatic ligation, if a "gap" exists, e polymerase and dNTPs may be added to "fill in" the gap). Eech primer has a portion thet ects as e "side chain" that does not bind the target sequence end acts es one helf of e stem structure that interects non-covalently through hydrogen bonding, salt bridges, van der Waal's forces, etc. Preferred embodiments utilize substantially complementary nucleic acids as the side

cheins. Thus, upon hybridization of the primers to the target sequence, the side chains of the primers are brought into spatial proximity, end, if the side chains comprise nucleic acids es well, can also form side chain hybridization complexes.

At least one of the side chains of the primers comprises an activatable cross-linking egent, generally covalently ettached to the side chain, thet upon activation, results in e chemical cross-link or chemical ligation. The ectivatible group may comprise any molety thet will ellow cross-linking of the side cheins, end include groups ectivated chemically, photonically and thermally, with photoactivatable groups being preferred. In some embodiments e single ectivatable group on one of the side chains is enough to result in cross-linking via interaction to e functional group on the other side chein; in alternate embodiments, activatable groups ere required on eech side cheln.

Once the hybridization complex is formed, and the cross-linking egent has been activated such that the primers have been covelently ettached, the reection is subjected to conditions to allow for the disessocation of the hybridization complex, thus freeing up the target to serve as a template for the next ligation or cross-linking. In this way, signal amplification occurs, and can be detected as outlined herein.

INVASIVE CLEAVAGE TECHNIQUES

In e preferred embodiment, the signel emplificetion technique is invasive cleavage technology, which is described in a number of patents end patent applications, including U.S. Petent Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; and 5,843,669, all of which are hereby incorporeted by reference in their entirety. Invasive cleavage technology is based on structure-specific nucleases that cleave nucleic ecids in a site-specific menner. Two probes are used: an "invader" probe and a "signalling" probe, thet edjacently hybridize to a target sequence with overlap. For mismatch discrimination, the Invader technology relies on complementarity et the overlap position where cleavaga occurs. The enzyme cleaves at the overlap, and releases the "tail" which may or may not be labeled. This can then be detected.

Generally, invasive cleavage technology may be described as follows. A target nucleic acid is recognized by two distinct probes. A first probe, generally referred to herein as an "invader" probe, is substantially complementary to a first portion of the target nucleic acid. A second probe, generally referred to herein as a "signal probe", is pertially complementery to the target nucleic ecid; the 3' end of tha signal oligonucleotide is substantially complementary to the terget sequence while the 5' end is non-complementary end preferably forms a single-stranded "tail" or "arm". The non-complementary end of the second probe preferably comprises e "generic" or "unique" sequence, frequently referred to herein as a "detection sequence", thet is used to indicete the presence or absence of the target nucleic ecid, as described below. The detection sequence of the second probe preferably comprises et leest one detectable label, although as outlined herein, since this detection sequence cen function as a

target sequence for e cepture probe, sandwich configurations utilizing lebel probes es described herein may elso be done.

Hybridizetion of the first and second oligonucleotides near or adjecent to one another on the target nucleic acid forms a number of structures. In a preferred embodiment, a forked cleavage structure forms and is a substrate of a nuclease which cleaves the detection sequence from the signel oligonucleotide. The site of cleavage is controlled by the distance or overlep between the 3' end of the invader oligonucleotide and the downstream fork of the signal oligonucleotide. Therefore, neither oligonucleotide is subject to cleavage when miseligned or when unattached to target nucleic acid.

In a preferred embodiment, the nuclease that recognizes the forked cleavage structure and catalyzes release of the tail is thermostable, thereby, allowing thermel cycling of the cleavage reaction, if desired. Praferred nucleeses derived from thermostable DNA polymerases that have been modified to have reduced synthetic activity which is an undesirable side-reaction during cleevage are disclosed in U.S. Patent Nos. 5,719,028 and 5,843,669, hereby expressly by reference. The synthetic activity of the DNA polymerase is reduced to a level where it does not interfere with detection of the cleavage reaction end detection of the freed tail. Preferably the DNA polymerase has no detectable polymerase activity. Examples of nucleeses ere those derived from *Thermus equaticus*, *Thermus flavus*, or *Thermus thermophilus*.

In enother embodiment, thermostable structure-specific nucleases are Flap endonucleases (FENs) selected from FEN-1 or FEN-2 lika (e.g. XPG and RAD2 nucleases) from Archaebacterial species, for example, FEN-1 from *Methenococcus jannaschii*, *Pyrococcus furiosis*, *Pyrococcus woesei*, and *Archaeoglobus fulgidus*. (U.S. Patent No. 5,843,669 and Lyamichev *et al.* 1999. Neture Biotechnology 17:292-297; both of which are hereby expressly by reference).

In a preferred embodiment, the nuclease is *Afu*FEN1 or *Pfu*FEN1 nuclease. To cleave a forked structure, these nucleases requira at least one overlapping nucleotide between the signal end invasive probes to recognize end cleave the 5' end of the signal probe. To effect cleavage the 3'-terminal nucleotide of the invader oligonucleotide is not required to be complementary to the target nucleic ecid. In contast, mismatch of the signal probe one base upstream of the cleavage site prevents creation of the overlap end cleavage. The specificity of the nuclease reaction allows single nucleotide polymorphism (SNP) detection from, for axample, genomic DNA, as outlined below (Lyamichev et al.).

The invasive cleavaga assay is preferably performed on an array format. In a preferred embodiment, the signal probe has a detectable label, ettached 5' from the site of nuclease cleavage (e.g. within the detection sequence) and e capture tag, as described below (e.g. biotin or other hapten) 3' from the site of nuclease cleavage. After the assay is carried out, the 3' portion of the cleeved signal probe (e.g. the

the detection sequence) ere extracted, for example, by binding to streptavidin beads or by crosslinking through the capture tag to produce aggregetes or by entibody to en attached hapten. By "capture tag" herein is e meant one of a pair of binding partners as described above, such es entigen/entibody pairs, digoxygenenin, dinitrophenol, etc.

The cleeved 5' region, e.g. the detection sequence, of the signal probe, comprises e lebel end is detected end optionally quentitated. In one embodiment, the cleaved 5' region is hybridized to e probe on an errey (capture probe) end optically detected. As described below, many signal probes cen be enelyzed in perallel by hybridization to their complementary probes in an array.

In a preferred embodiment, the invasive cleavage reaction is configured to utilize e fluorophore-quencher reection. A signalling probe comprising both e fluorophore and a quencher is used, with the fluorophore end the quencher on opposite sides of the cleevage site. As will be appreciated by those in the ert, these will be positioned closely together. Thus, in the absence of cleavage, very little signal is seen due to the quenching reaction. After cleavage, however, the distance between the two is large, and thus fluorescence can be detected. Upon assembly of an essay complex, comprising the target sequence, an invader probe, and a signalling probe, and the introduction of the cleavage enzyme, the cleavage of the complex results in the disessociation of the quencher from the complex, resulting in an increase in fluorescence.

In this embodiment, suitable fluorophore-quencher pairs ere es known in the art. For example, suitable quencher molecules comprise Dabcyl.

As will be appreciated by those in the art, this system can be configured in a variety of conformations, as discussed in Figure 4.

In a preferred embodiment, to obtain higher specificity end reduce the detection of contaminating uncleaved signal probe or incorrectly cleaved product, en eddltionel enzymatic recognition step is introduced in the erray capture procedure. For example, the cleaved signal probe binds to a capture probe to produce e double-strended nucleic acid in the array. In this embodiment, the 3' end of the cleaved signal probe is edjacent to the 5' end of one strend of the capture probe, thereby, forming a substrate for DNA ligase (Broude et al. 1991. PNAS 91: 3072-3076). Only correctly cleaved product is ligeted to the capture probe. Other incorrectly hybridized and non-cleaved signal probes are removed, for example, by heat deneturation, high stringency washes, end other methods that disrupt base pairing.

CYCLING PROBE TECHNIQUES (CPT)

In a preferred embodiment, the signal amplification technique is CPT. CPT technology is described in e number of petents and patent applications, including U.S. Patent Nos. 5,011,769, 5,403,711,

5,660,988, and 4,876,187, and PCT published applications WO 95/05480, WO 95/1416, and WO 95/00667, end U.S.S.N. 09/014,304, all of which ere expressly incorporated by referenca in their entirety.

Generally, CPT may be described es follows. A CPT primer (also sometimes referred to herein es a "scissile primer"), comprises two probe sequences separated by e scissile linkage. The CPT primer is substantially complementary to the target saquence end thus will hybridize to it to form a hybridization complex. The scissile linkage is cleaved, without cleaving the target sequence, resulting in the two probe sequences being separated. The two probe sequences can thus be more easily disassociated from the target, and the reection can be repeated any number of times. Tha cleaved primer is then detected as outlined herein.

By "scissile linkege" herein is meant a linkage within the scissila probe that can be cleaved when the probe is part of e hybridization complex, thet is, when a double-stranded complex is formed. It is important that the scissila linkage cleave only the scissile probe and not the sequence to which it is hybridized (i.e. aither the target sequence or a probe sequence), such that the target sequence may be reused in the reaction for amplification of the signal. As used herein, the scissile linkage, is eny connecting chemical structure which joins two probe sequences end which is capable of being selectively cleaved without cleavage of either the probe sequences or the saquence to which the scissile proba is hybridizad. The scissile linkage may be a single bond, or a multiple unit sequence. As will be appreciated by those in the ert, a number of possible scissile linkages may be used.

In a preferred embodiment, the scissile linkage comprises RNA. This system, previously described in as outlined ebove, is based on the fact that certain double-stranded nucleases, particularly ribonucleases, will nick or excise RNA nucleosides from a RNA:DNA hybridization complex. Of particular use in this embodiment is RNAseH, Exo III, and reverse transcriptase.

In one embodiment, the antire scissile probe is made of RNA, the nicking is facilitated especially when carried out with a double-stranded ribonuclease, such as RNAseH or Exo III. RNA probes made entirely of RNA sequences are particularly useful because first, they can be more easily produced enzymatically, and second, they have more cleavage sites which are accessible to nicking or cleaving by a nicking agent, such as the ribonucleases. Thus, scissile probes made entirely of RNA do not rely on a scissile linkage since the scissile linkage is inherent in the probe.

In a preferred embodiment, when the scissile linkage is a nucleic ecid such as RNA, the methods of the invention may be used to detect mismatches, as is generally described in U.S. Patent Nos. 5,660,988, and WO 95/14106, hereby expressly incorporated by reference. These mismatch detection methods are based on the fact that RNAseH may not bind to and/or cleave an RNA:DNA duplex if there are mismatches present in the sequence. Thus, in the NA_1-R-NA_2 embodiments, NA_1

and NA_2 are non-RNA nucleic acids, preferably DNA. Preferably, the mismatch is within the RNA:DNA duplex, but in some embodiments the mismatch is present in an adjacent sequence very close to the desired sequence, close enough to affect the RNAseH (generally within one or two beses). Thus, in this embodiment, the nucleic acid scissile linkage is designed such that the sequence of the scissile linkage reflects the particular sequence to be detected, i.e. the aree of the putative mismatch.

In some embodiments of mismatch detection, the rate of generetion of the released fragments is such that the methods provide, essentielly, e yes/no result, whereby the detection of virtually eny releesed fragment indicetes the presence of the desired terget sequence. Typically, however, when there is only e minimel mismatch (for example, e 1-, 2- or 3-base mismatch, or a 3-bese deletion), there is some generation of cleaved sequences even though the target sequence is not present. Thus, the rate of generetion of cleeved fragments, end/or the final amount of cleaved fragments, is quantified to indicete the presence or absence of the target. In eddition, the use of secondary and tertiary scissile probes may be particularly useful in this embodiment, as this can amplify the differences between a perfect match and a mismatch. These methods may be particularly useful in the determination of homozygotic *or* heterozygotic states of a patient.

In this embodiment, it is en important feature of the scissile linkage that its length is determined by the suspected difference between the target end the probe. In particular, this means that the scissile linkege must be of sufficient length to encompass the suspected difference, yet short enough so thet the scissile linkage cannot inappropriately "specifically hybridize" to the selected nucleic ecid molecule when the suspected difference is present; such inappropriate hybridizetion would permit excision and thus cleavage of scissile linkages even though the selected nucleic acid molecule was not fully complementary to the nucleic acid probe. Thus in a preferred embodiment, the scissile linkege is between 3 to 5 nucleotides in length, such that a suspected nucleotide difference from 1 nucteotide to 3 nucleotides is encompassed by the scissile linkage, end 0, 1 or 2 nucleotides are on either side of the difference.

Thus, when the scissile linkage is nucleic acid, preferred embodiments utilize from 1 to about 100 nucleotides, with from about 2 to ebout 20 being preferred and from about 5 to about 10 being particularly preferred.

CPT may be done enzymatically or chemically. That is, in eddition to RNAseH, there are several other cleaving agents which mey be useful in cleaving RNA (or other nucleic ecid) scissile bonds. For example, several chemical nucleases heve been reported; see for exemple Sigmen et al., Annu. Rev. Biochem. 1990, 59, 207-236; Sigman et al., Chem. Rev. 1993, 93, 2295-2316; Bashkin et al., J. Org. Chem. 1990, 55, 5125-5132; end Sigman et al., Nucleic Acids end Moleculer Biology, vol. 3, F. Eckstein and D.M.J. Lilley (Eds), Springer-Verlag, Heidelberg 1989, pp. 13-27; all of which are hereby expressly incorporated by reference.

Specific RNA hydrolysis is also an active area; see for example Chin, Acc. Chem. Res. 1991, 24, 145-152; Breslow et al., Tetrahedron, 1991, 47, 2365-2376; Anslyn et al., Angew. Chem. Int. Ed. Engl., 1997, 36, 432-450; and references therein, all of which are expressly incorporated by reference. Reactive phosphate centers are also of interest in developing scissile linkages, see Hendry et al., Prog. Inorg. Chem. : Bioinorganic Chem. 1990, 31, 201-258 also expressly incorporated by reference.

Current approaches to site-directed RNA hydrolysis include the conjugation of a reactive moiety capable of cleaving phosphodiester bonds to a recognition element capable of sequence-specifically hybridizing to RNA. In most cases, a metal complex is covalently attached to a DNA strand which forms a stable heteroduplex. Upon hybridization, a Lewis acid is placed in close proximity to the RNA backbone to effect hydrolysis; see Magda et al., J. Am. Chem. Soc. 1994, 116, 7439; Hall et al., Chem. Biology 1994, 1, 185-190; Bashkin et al., J. Am. Chem. Soc. 1994, 116, 5981-5982; Hall et al., Nucleic Acids Res. 1996, 24, 3522; Magda et al., J. Am. Chem. Soc. 1997, 119, 2293; and Magda et al., J. Am. Chem. Soc. 1997, 119, 6947, all of which are expressly incorporated by reference.

In a similar fashion, DNA-polyamine conjugates have been demonstrated to induce site-directed RNA strand scission; see for example, Yoshinari et al., J. Am. Chem. Soc. 1991, 113, 5899-5901; Endo et al., J. Org. Chem. 1997, 62, 846; and Barbier et al., J. Am. Chem. Soc. 1992, 114, 3511-3515, all of which are expressly incorporated by reference.

In a preferred embodiment, the scissile linkage is not necessarily RNA. For example, chemical cleavage moieties may be used to cleave basic sites in nucleic acids; see Belmont, et al.,New J. Chem. 1997, 21, 47-54; and references therein, all of which are expressly incorporated herein by reference. Similarly, photocleavable moieties, for example, using transition metals, may be used; see Moucheron, et al., Inorg. Chem. 1997, 36, 584-592, hereby expressly by reference.

Other approaches rely on chemical moieties or enzymes; see for example Keck et al., Biochemistry 1995, 34, 12029-12037; Kirk et al., Chem. Commun. 1998, in press; cleavage of G-U basepairs by metal complexes; see Biochemistry, 1992, 31, 5423-5429; diamine complexes for cleavage of RNA; Komiyama, et al., J. Org. Chem. 1997, 62, 2155-2160; and Chow et al., Chem. Rev. 1997, 97, 1489-1513, and references therein, all of which are expressly incorporated herein by reference.

The first step of the CPT method requires hybridizing a primary scissile primer (also called a primary scissile probe) to the target. This is preferably done at a temperature that allows both the binding of the longer primary probe and disassociation of the shorter cleaved portions of the primary probe, as will be appreciated by those in the art. As outlined herein, this may be done in solution, or either the target or one or more of the scissile probes may be attached to a solid support. For example, it is possible to utilize "anchor probes" on a solid support which are substantially complementary to a portion of the target sequence, preferably a sequence that is not the same sequence to which a

scissile probe will bind.

Similarly, as outlined herein, a preferred embodiment has one or more of the scissile probes attached to a solid support such as a bead. In this embodiment, the soluble target diffuses to allow the formation of the hybridization complex between the soluble target sequence and the support-bound scissile probe. In this embodiment, it may be desirable to include additional scissile linkages in the scissile probes to allow the release of two or more probe sequences, such that more than one probe sequence per scissile probe may be detected, as is outlined below, in the interests of maximizing the signal.

In this embodiment (and in other techniques herein), preferred methods utilize cutting or shearing techniques to cut the nucleic acid sample containing the target sequence into a size that will allow sufficient diffusion of the target sequence to the surface of a bead. This may be accomplished by shearing the nucleic acid through mechanical forces (e.g. sonication) or by cleaving the nucleic acid using restriction endonucleases. Alternatively, a fragment containing the target may be generated using polymerase, primers and the sample as a template, as in polymerase chain reaction (PCR). In addition, amplification of the target using PCR or LCR or related methods may also be done; this may be particularly useful when the target sequence is present in the sample at extremely low copy numbers. Similarly, numerous techniques are known in the art to increase the rate of mixing and hybridization including agitation, heating, techniques that increase the overall concentration such as precipitation, drying, dialysis, centrifugation, electrophoresis, magnetic bead concentration, etc.

In general, the scissile probes are introduced in a molar excess to their targets (including both the target sequence or other scissile probes, for example when secondary or tertiary scissile probes are used), with ratios of scissile probe:target of at least about 100:1 being preferred, at least about 1000:1 being particularly preferred, and at least about 10,000:1 being especially preferred. In some embodiments the excess of probe:target will be much greater. In addition, ratios such as these may be used for all the amplification techniques outlined herein.

Once the hybridization complex between the primary scissile probe and the target has been formed, the complex is subjected to cleavage conditions. As will be appreciated, this depends on the composition of the scissile probe; if it is RNA, RNAseH is introduced. It should be noted that under certain circumstances, such as is generally outlined in WO 95/00666 and WO 95/00667, hereby incorporated by reference, the use of a double-stranded binding agent such as RNAseH may allow the reaction to proceed even at temperatures above the Tm of the primary probe:target hybridization complex. Accordingly, the addition of scissile probe to the target can be done either first, and then the cleavage agent or cleavage conditions introduced, or the probes may be added in the presence of the cleavage agent or conditions.

The cleavage conditions result in the separation of the two (or more) probe sequences of the primery scissile probe. As a result, the shorter probe sequences will no longer remain hybridized to the target sequence, end thus the hybridizetion complex will disassociete, leaving the target sequence intect.

The optimal tempereture for cerrying out the CPT reactions is genarelly from ebout 5°C to about 25°C below the melting temperatures of the probe:target hybridization complex. This provides for e rapid rate of hybridization and high degree of specificity for the target sequence. The Tm of eny particuler hybridization complex depends on salt concentretion, G-C content, end length of the complex, as is known in the art end described herein.

During tha reaction, as for the other amplification techniques herein, it may be nacessary to suppress cleavage of the probe, es well es the target sequence, by nonspecific nucleases. Such nucleases ere generally removed from tha sample during the isolation of the DNA by heating or extraction procedures. A number of inhibitors of single-strended nucleases such es vanadate, inhibitors it-ACE and RNAsin, a placantal protain, do not affect the activity of RNAsaH. This may not be necessary depending on the purity of the RNAseH and/or the target semple.

These steps ere repeated by allowing the reaction to proceed for a period of time. The reaction is usually carried out for about 15 minutes to about 1 hour. Generally, each molecule of the target sequence will turnover between 100 and 1000 times in this period, depending on the length and sequence of the probe, the specific reection conditions, and the cleavage method. For example, for each copy of the target sequence present in the test semple 100 to 1000 molecules will be cleaved by RNAseH. Higher levels of emplification can be obtained by allowing the reaction to proceed longer, or using secondery, tertiary, or quaternary probes, as is outlined herein.

Upon completion of the reaction, generally determined by time or amount of cleavage, the uncleeved scissile probes must be removed or neutralized prior to detection, such that the uncleaved probe does not bind to e detection probe, causing false positive signals. This may be done in a variety of ways, as is generally described below.

In a preferred embodiment, the separetion is facilitated by the use of beads contalning the primary probe. Thus, when tha scissila probes are attached to beads, ramoval of the beads by filtration, centrifugation, the application of a megnetic field, electrostatic interactions for charged beeds, adhesion, etc., results in the removal of the uncleaved probes.

In a preferred embodiment, the separation is based on strong acid precipitation. This is useful to separate long (generally greater than 50 nucleotides) from smaller fregments (generelly ebout 10 nucleotides). The introduction of e strong acid such es trichloroacetic acid into the solution causes the longer probe to precipitate, while the smaller cleaved fragments remeln in solution. The solution can

be centrifuged or filtered to remove the precipitate, and the cleaved probe sequences can be quantitated.

In a preferred embodiment, the scissile probe contains both a detectable label end an affinity binding ligand or moiety, such that en affinity support is used to carry out the separation. In this embodiment, it is important that the detectable label used for detection is not on the same probe sequence that contains the affinity moiety, such that removal of the uncleaved probe, and the cleaved probe contalning the affinity molety, does not remove all the detectable labels. Alternatively, the scissile probe may contain a capture tag; the binding partner of the capture tag is attached to a solid support such as glass beeds, latex beads, dextrans, etc. and used to pull out the uncleaved probes, as is known in the art. The cleeved probe sequences, which do not contain the capture tag, remain in solution and then can be detected as outlined below.

In a preferred embodiment, simllar to the above embodiment, e separation sequence of nucleic acid is included In the scissile probe, which is not cleaved during the reaction. A nucleic acid complementary to the separation sequence is attached to a solid support such as a bead and serves as a catcher sequence. Preferably, the separation sequence is added to the scissile probes, and is not recognized by the target sequence, such that a generalized catcher sequence may be utilized in a variety of assays.

After removal of the uncleaved probe, es required, detection proceeds via the addition of the cleaved probe sequences to the erray compositions, as outlined below. In general, the cleaved probe is bound to a capture probe, either directly or indirectly, and the label is detected. In a preferred embodiment, no higher order probes are used, end detection is based on the probe sequence(s) of the primery primer. In a preferred embodiment, at least one, end preferably more, secondary probes (also referred to herein as secondary primers) are used; the secondary probes hybridize to the domains of the cleavage probes; etc.

Thus, CPT requires, egain in no particular order, a first CPT primer comprising a first probe sequence, a scissile linkage and a second probe sequence; and a cleavage agent.

In this manner, CPT results in the generation of a large amount of cleaved primers, which then can be detected as outlined below.

SANDWICH ASSAY TECHNIQUES

In a preferred embodiment, the signal emplification technique is a "sandwich" assay, as is generally described in U.S.S.N. 60/073,011 end In U.S. Patent Nos. 5,681,702, 5,597,909, 5,545,730, 5,594,117, 5,591,584, 5,571,670, 5,580,731, 5,571,670, 5,591,584, 5,624,802, 5,635,352, 5,594,118, 5,359,100, 5,124,246 end 5,681,697, all of whlch ere hereby incorporated by reference. Although

sandwich assays do not result in the elteretion of primers, sendwich essays can be considered signal emplification techniques since multiple signels (i.e. label probes) are bound to e single target, resulting in the emplificetion of the signal. Sandwich essays mey be used when the target sequence does not contain e label; or when adapters are used, as outlined below.

As discussed herein, it should be noted thet the sandwich assays can be used for the detection of primary target sequences (e.g. from a patient sample), or es e method to detect the product of en amplification reaction as outlined ebove; thus for example, eny of the newly synthesized strands outlined ebove, for example using PCR, LCR, NASBA, SDA, etc., may be used es the "target sequence" in a sandwich assey.

As will be appreciated by those in the art, the systems of the invention mey teke on a large number of different configurations. In general, there ere three types of systems that can be used: (1) "non-sendwich" systems (also referred to herein es "direct" detection) in which the target sequence itself is labeled with detectable lebels (again, either because the primers comprise labels or due to the incorporation of labels into the newly synthesized strend); (2) systems in which label probes directly bind to the target sequences; and (3) systems in which label probes are indirectly bound to the target sequences, for example through the use of amplifier probes.

The enchoring of the target sequence to the beed is done through the use of capture probes and optionally either capture extender probes (sometimes referred to as "adapter sequences" herein). When only capture probes ere utilized, it is necessary to have unique cepture probes for each target sequence; thet is, the surface must be customized to contain unique capture probes; e.g. each bead comprises a different capture probe. Alternatively, capture extender probes mey be used, that allow e "universal" surface, i.e. e surfece containing a single type of capture probe thet cen be used to detect any target sequence. "Capture extender" probes have a first portion thet will hybridize to all or part of the capture probe, and a second portion that will hybridize to a first portion of the target sequence. This then allows the generetion of customized soluble probes, which es will be eppreciated by those in the art is generally simpler and less costly. As shown herein, two capture extender probes may be used. This has generally been done to stabilize essay complexes for example when the target sequence is large, or when lerge emplifier probes (particularly branched or dendrimer amplifier probes) are used.

Detection of the amplification reactions of the invention, including the direct detection of emplification products end indirect detection utilizing label probes (i.e. sandwich essays), is preferably done by detecting essay complexes comprising detectable labels, which cen be attached to the assay complex in a variety of ways, as is more fully described below.

Once the target sequence has preferebly been anchored to the array, an emplifier probe is hybridized

to the target sequence, either directly, or through the use of one or more label extender probes, which serves to allow "ganeric" amplifiar probes to be mede. As for ell tha steps outlined herein, this may be done simultaneously with capturing, or sequantially. Prefarably, the amplifier probe contains a multiplicity of emplificetion sequences, although in some ambodiments, as dascribed below, tha amplifier probe may contain only a single amplification sequence, or at least two amplification sequences. The amplifier probe may take on a number of different forms; either a branched conformation, a dendrimer conformation, or e linear "string" of amplification sequences. Label probes comprising detectable labels (preferably but not raquired to be fluorophores) then hybridize to the emplification sequences (or in some cases the label probes hybridize directly to the target sequence), and tha labels detectad, as is more fully outlined below.

Accordingly, the present invention provides compositions comprising an amplifier probe. By "amplifier probe" or "nucleic acid multimer" or "amplificetion multimar" or grammatical equivalents herain is meant a nucleic acid probe that is used to facilitate signal amplification. Amplifier probes comprise at least a first single-stranded nucleic acid probe sequence, as defined below, and at least one single-stranded nucleic acid amplification sequence, with e multiplicity of emplification sequences being praferred.

Amplifier probes comprise a first proba sequence that is used, either directly or indirectly, to hybridize to the target sequence. That is, the amplifier probe itself may have a first probe sequence that is substantially complementary to the target sequance, or it has a first probe sequence that is substantially complementary to a portion of an additional probe, in this case called a label extender probe, that has a first portion that is substantially complementary to the target sequence. In a preferred embodiment, the first probe sequence of the amplifier probe is substantially complementary to the target sequence.

In general, as for all the probes herein, the first probe sequence is of a length sufficient to give specificity and stability. Thus generally, the probe sequences of the invention that are designed to hybridize to another nucleic acid (i.e. probe sequences, amplification sequences, portions or domains of larger probes) ara et least about 5 nucleosides long, with at least about 10 being preferred and at least about 15 being especially preferred.

In a preferrad embodiment, severel different amplifier probes ere used, aach with first probe sequences that will hybridize to a different portion of the target sequence. That is, thera is more than one level of amplification; the amplifier probe provides an amplification of signal due to a multiplicity of labelling evants, and saveral different emplifier probes, each with this multiplicity of labels, for each target sequence is used. Thus, preferred embodiments utilize at least two different pools of emplifier probes, each pool having e different probe sequence for hybridization to different portions of the target sequence; the only real limitation on the number of different amplifier probes will be the length of the

original target sequence. In addition, it is also possible that the different amplifier probes contain different amplification sequences, although this is generally not preferred.

In a preferred embodiment, the amplifier probe does not hybridize to the sample target sequence directly, but instead hybridizes to a first portion of a label extender probe. This is particularly useful to allow the use of "generic" amplifier probes, that is, amplifier probes that can be used with a variety of different targets. This may be desirable since several of the amplifier probes require special synthesis techniques. Thus, the addition of a relatively short probe as a label extender probe is preferred. Thus, the first probe sequence of the amplifier probe is substantially complementary to a first portion or domain of a first label extender single-stranded nucleic acid probe. The label extender probe also contains a second portion or domain that is substantially complementary to a portion of the target sequence. Both of these portions are preferably at least about 10 to about 50 nucleotides in length, with a range of about 15 to about 30 being preferred. The terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target or probe sequences. For example, assuming a 5'-3' orientation of the complementary target sequence, the first portion may be located either 5' to the second portion, or 3' to the second portion. For convenience herein, the order of probe sequences are generally shown from left to right.

In a preferred embodiment, more than one label extender probe-amplifier probe pair may be used, that is, n is more than 1. That is, a plurality of label extender probes may be used, each with a portion that is substantially complementary to a different portion of the target sequence; this can serve as another level of amplification. Thus, a preferred embodiment utilizes pools of at least two label extender probes, with the upper limit being set by the length of the target sequence.

In a preferred embodiment, more than one label extender probe is used with a single amplifier probe to reduce non-specific binding, as is generally outlined in U.S. Patent No. 5,681,697, incorporated by reference herein. In this embodiment, a first portion of the first label extender probe hybridizes to a first portion of the target sequence, and the second portion of the first label extender probe hybridizes to a first probe sequence of the amplifier probe. A first portion of the second label extender probe hybridizes to a second portion of the target sequence, and the second portion of the second label extender probe hybridizes to a second probe sequence of the amplifier probe. These form structures sometimes referred to as "cruciform" structures or configurations, and are generally done to confer stability when large branched or dendrimeric amplifier probes are used.

In addition, as will be appreciated by those in the art, the label extender probes may interact with a preamplifier probe, described below, rather than the amplifier probe directly.

Similarly, as outlined above, a preferred embodiment utilizes several different amplifier probes, each with first probe sequences that will hybridize to a different portion of the label extender probe. In

eddition, as outlined above, it is elso possible that the different amplifier probes contain different emplification sequences, elthough this is generally not preferred.

In addition to the first probe sequence, the amplifier probe also comprises et leest one emplificetion sequence. An "amplification sequence" or "amplification segment" or grammaticel equivalents herein is meant e sequence thet is used, either directly or indirectly, to bind to a first portion of a lebel probe es is more fully described below. Preferebly, the amplifier probe comprises a multiplicity of emplification aequences, with from ebout 3 to ebout 1000 being preferred, from about 10 to about 100 being perticularly preferred, end about 50 being especially preferred. In some cases, for example when linear amplifier probes are used, from 1 to about 20 is preferred with from about 5 to about 10 being particularly preferred.

The amplification sequences may be linked to each other in a variety of ways, es will be eppreciated by those in the ert. They may be covalently linked directly to eech other, or to intervening sequences or chemical moieties, through nucleic acid linkeges such as phosphodiester bonds, PNA bonds, etc., or through interposed linking agents such emino acid, carbohydrate or polyol bridges, or through other cross-linking agents or binding partners. The site(s) of linkage may be et the ends of a segment, and/or at one or more internal nucleotides in the strand. In e preferred embodiment, the amplification sequences ere ettached via nucleic acid linkages.

In a preferred embodiment, brenched amplifier probes are used, es are generally described in U.S. Patent No. 5,124,246, hereby incorporated by reference. Branched amplifier probes may take on "fork-like" or "comb-like" conformations. "Fork-like" branched amplifier probes generally have three or more oligonucleotide segments emanating from a point of origin to form a branched structure. The point of origin may be another nucleotide segment or a multifunctional molecule to whcih et least three segments can be covalently or tightly bound. "Comb-like" branched amplifier probes have a linear backbone with a multiplicity of sidechain oligonucleotides extending from the backbone. In either conformation, the pendent segments will normally depend from a modified nucleotide or other organic molety having the appropriate functional groups for attachment of oligonucleotides. Furthermore, in either conformation, e large number of emplification sequences are available for binding, either directly or indirectly, to detection probes. In general, these structures ere made as is known in the art, using modified multifunctional nucleotides, as is described in U.S. Patent Nos. 5,635,352 end 5,124,246, among others.

In e preferred embodiment, dendrimer amplifier probes ere used, as are generally described in U.S. Patent No. 5,175,270, hereby expressly incorporated by reference. Dendrimeric emplifier probes have emplification sequences that ere attached via hybridizetion, and thus heve portions of double-stranded nucleic acid as e component of their structure. The outer surface of the dendrimer amplifier probe has a multiplicity of amplification sequences.

In a preferred embodiment, linear amplifier probes are used, that have individual amplification sequences linked end-to-end either directly or with short intervening sequences to form a polymer. As with the other amplifier configurations, there may be additional sequences or moieties between the amplification sequences. In one embodiment, the linear amplifier probe has a single amplification sequence.

In addition, the amplifier probe may be totally linear, totally branched, totally dendrimeric, or any combination thereof.

The amplification sequences of the amplifier probe are used, either directly or indirectly, to bind to a label probe to allow detection. In a preferred embodiment, the amplification sequences of the amplifier probe are substantially complementary to a first portion of a label probe. Alternatively, amplifier extender probes are used, that have a first portion that binds to the amplification sequence and a second portion that binds to the first portion of the label probe.

In addition, the compositions of the invention may include "preamplifier" molecules, which serves a bridging moiety between the label extender molecules and the amplifier probes. In this way, more amplifier and thus more labels are ultimately bound to the detection probes. Preamplifier molecules may be either linear or branched, and typically contain in the range of about 30-3000 nucleotides.

Thus, label probes are either substantially complementary to an amplification sequence or to a portion of the target sequence.

Detection of the amplification reactions of the invention, including the direct detection of amplification products and indirect detection utilizing label probes (i.e. sandwich assays), is done by detecting assay complexes comprising labels as is outlined herein.

In addition to amplification techniques, the present invention also provides a variety of genotyping reactions that can be similarly detected and/or quantified.

GENOTYPING

In this embodiment, the invention provides compositions and methods for the detection (and optionally quantification) of differences or variations of sequences (e.g. SNPs) using bead arrays for detection of the differences. That is, the bead array serves as a platform on which a variety of techniques may be used to elucidate the nucleotide at the position of interest ("the detection position"). In general, the methods described herein relate to the detection of nucleotide substitutions, although as will be appreciated by those in the art, deletions, insertions, inversions, etc. may also be detected.

These techniques fall into five general categories: (1) techniques that rely on traditional hybridization

methods that utilize the variation of stringency conditions (temperature, buffer conditions, etc.) to distinguish nucleotides at the detection position; (2) extension techniques that add a base ("the base") to basepair with the nucleotide at the detection position; (3) ligation techniques, that rely on the specificity of ligase enzymes (or, in some cases, on the specificity of chemical techniques), such that ligation reactions occur preferentially if perfect complementarity exists at the detection position; (4) cleavage techniques, that also rely on enzymatic or chemical specificity such that cleavage occurs preferentially if perfect complementarity exists; and (5) techniques that combine these methods.

As outlined herein, in this embodiment the target sequence comprises a position for which sequence information is desired, generally referred to herein as the "detection position" or "detection locus". In a preferred embodiment, the detection position is a single nucleotide, although in some embodiments, it may comprise a plurality of nucleotides, either contiguous with each other or separated by one or more nucleotides. By "plurality" as used herein is meant at least two. As used herein, the base which basepairs with a detection position base in a hybrid is termed a "readout position" or an "interrogation position".

In some embodiments, as is outlined herein, the target sequence may not be the sample target sequence but instead is a product of a reaction herein, sometimes referred to herein as a "secondary" or "derivative" target sequence. Thus, for example, in SBE, the extended primer may serve as the target sequence; similarly, in invasive cleavage variations, the cleaved detection sequence may serve as the target sequence.

As above, if required, the target sequence is prepared using known techniques. Once prepared, the target sequence can be used in a variety of reactions for a variety of reasons. For example, in a preferred embodiment, genotyping reactions are done. Similarly, these reactions can also be used to detect the presence or absence of a target sequence. In addition, in any reaction, quantitation of the amount of a target sequence may be done. While the discussion below focuses on genotyping reactions, the discussion applies equally to detecting the presence of target sequences and/or their quantification.

Furthermore, as outlined below for each reaction, each of these techniques may be used in a solution based assay, wherein the reaction is done in solution and a reaction product is bound to the array for subsequent detection, or in solid phase assays, where the reaction occurs on the surface and is detected.

These reactions are generally classified into 5 basic categories, as outlined below.

SIMPLE HYBRIDIZATION GENOTYPING

In a preferred embodiment, straight hybridization methods are used to elucidate the identity of the

base at the detection position. Generally speaking, these techniques break down into two basic types of reactions: those that rely on competitive hybridization techniques, and those that discriminate using stringency parameters and combinations thereof.

Competitive hybridization

In a preferred embodiment, the use of competitive hybridization probes is done to elucidate either the identity of the nucleotide(s) at the detection position or the presence of a mismatch. For example, sequencing by hybridization has been described (Drmanac et al., Genomics 4:114 (1989); Koster et el., Nature Biotechnology 14:1123 (1996); U.S. Patent Nos. 5,525,464; 5,202,231 and 5,695,940, among others, ell of which are hereby expressly incorporated by reference in their entirety).

It should be noted in this context that "mismatch" is a relative term and meant to indicate a difference in the identity of a base at a particular position, termed the "detection position" herein, between two sequences. In general, sequences that differ from wild type sequences are referred to as mismatches. However, particularly in the case of SNPs, what constitutes "wild type" may be difficult to determine as multiple alleles can be relatively frequently observed in the population, and thus "mismatch" in this context requires the artificial adoption of one sequence es a standerd. Thus, for the purposes of this invention, sequences are referred to herein as "match" and "mismatch". Thus, the present invention may be used to detect substitutions, insertions or deletions as compared to a wild-type sequence.

In a preferred embodiment, a plurality of probes (sometimes referred to herein as "readout probes") are used to identify the base at the detection position. In this embodiment, each different readout probe comprises a different detection label (which, as outlined below, can be either a primary label or a secondary label) and a different base at the position that will hybridize to the detection position of the target sequence (herein referred to as the readout position) such that differential hybridization will occur. That is, all other parameters being equal, e perfectly complementary readout probe (a "match probe") will in general be more stable and have a slower off rate than a probe comprising a mismatch (a "mismatch probe") at any particular temperature. Accordingly, by using different readout probes, each with a different base at the readout position and each with a different label, the identification of the base at the detection position is elucidated.

Accordingly, a detectable label is incorporated into the readout probe. In a preferred embodiment, a set of readout probes are used, each comprising a different base at the readout position. In some embodiments, each readout probe comprises a different label, that is distinguishable from the others. For example, a first label may be used for probes comprising adenosine at the readout position, a second label may be used for probes comprising guanine at the readout position, etc. In a preferred embodiment, the length and sequence of each readout probe is identical except for the readout position, although this need not be true in all embodiments.

The number of reedout probes used will vary depending on the end use of the essay. For exemple, many SNPs are blallelic, end thus two readout probes, each comprising en interrogetion bese thet will basepair with one of the detection position beses. For sequencing, for example, for the discovery of SNPs, a set of four readout probes ere used, although SNPs may elso be discovered with fewer reedout perameters.

As will be appreciated by those in the ert end edditionally outlined below, this system can take on e number of different configuretions, including e solution phase essay end e solid phese assey.

Solution phase assey

A solution phase assay that is followed by ettaching the target sequence to an erray is depicted in Figure 8D. In Figure 8D, a reaction with two different readout probes is shown. After the competitive hybridization has occured, the target sequence is added to the erray, which may take on severel configurations, outlined below.

Solid phase essay

In e preferred embodiment, the competition reaction is done on the array. This system may take on several configurations.

In e preferred embodiment, e sandwich essay of sorts is used. In this embodiment, the beed comprises a capture probe that will hybridize to a first target domain of a target sequence, and the readout probe will hybridize to a second target domein, as is generally depicted in Figure 8A. In this embodiment, the first target domain may be either unique to the target, or may be en exogeneous adapter sequence edded to the terget sequence as outlined below, for example through the use of PCR reactions. Simllarly, a sandwich assay that utilizes e cepture extender probe, es described below, to ettach the target sequence to the array is depicted in Figure 8C.

Alternatively, the capture probe itself can be the readout probe as is shown in Figure 8B; that is, e plurality of microspheres are used, each comprising a capture probe that has e different base at the readout position. In general, the target sequence then hybridizes preferentially to the ceptura probe most closely matched. In this embodiment, either the target sequence itself is lebeled (for example, it may be the product of an amplification reaction) or a label probe may bind to the target sequence at a domeln remote from the detection position. In this embodiment, since it is the locetion on the erray thet serves to identify the bese et the detection position, different labels ere not required.

In e further embodiment, the target sequence itself is attached to the array, as generelly depicted for bead erreys in Figure 8E and described below.

Stringency Variation

In a preferred embodiment, sensitivity to variations in stringency parameters are used to determine either the identity of the nucleotide(s) et the detection position or the presence of e mismetch. As e preliminery metter, the use of different stringency conditions such es varietions in temperature end buffer composition to determine the presence or ebsence of mismatches in double stranded hybrids comprising e single strended target saquence and a probe is well known.

With particular regerd to temperature, es is known in the ert, differences in the number of hydrogen bonds as a function of basepairing between perfect matches and mismatches cen be exploited as e result of their different Tms (the tempereture at which 50% of the hybrid is denatured). Accordingly, a hybrid comprising perfect complementarity will melt at a higher temperature than one comprising et leest one mismetch, ell other parameters being equel. (It should be noted that for the purposes of tha discussion herein, all other parameters (i.e. length of the hybrid, nature of the backbone (i.e. naturelly occuring or nucleic ecid anatog), the assay solution composition and the composition of the bases, including G-C content are kept constant). However, as will be appreciated by those in the art, thase factors may be varied es well, end then teken into account.)

In general, as outlined herein, high stringency conditions ere those that result in perfect matches remaining in hybridization complexes, while imperfect matches melt off. Similarly, low stringency conditions are those thet ellow the formetion of hybridization complexes with both perfect and imperfect matches. High stringency conditions are known in the art; see for example Maniatis et el., Molecular Cloning: A Laboretory Menuel, 2d Edition, 1989, end Short Protocols in Moleculer Biology, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions ere sequence-dependent and will be different in different circumstances. Longer sequances hybridize specificelly at higher temperatures. An extensive guide to the hybridization of nucleic ecids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization end the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be ebout 5-10°C lower than the thermal melting point (T_m) for the specific sequence et a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH and nucleic ecid concentration) at which 50% of the probes complementary to the target hybridize to the terget sequence at equilibrium (as the target sequences ere present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less then ebout 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g. 10 to 50 nucleotides) and at least ebout 60°C for long probes (e.g. greater then 50 nucleotides). Stringent conditions mey elso be achieved with the eddition of destabilizing agents such es formamide. In enother embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, es ere known in the ert; see Manietis end Ausubel, supra, and Tijssen, supre.

As will be appreciated by those in the ert, mismetch detection using temperature mey proceed in a variety of ways, and is similar to the use of readout probes as outlined above. Again, es oultined above, a plurality of readout probes may be used in a sandwich format; in this embodiment, all the probes may bind at permissive, low temperatures (temperetures below the Tm of the mismatch); however, repeating the assay et a higher temperature (above the Tm of the mismatch) only the perfectly matched probe may bind. Thus, this system may be run with readout probes with differant detectable labels, es outlined above. Alternetively, a single probe mey be used to query whather e particular base is present.

Alternatively, es described above, the capture probe may serve as the readout probe; in this embodiment, a single label may be used on the target; at temperatures above the Tm of the mismatch, only signals from perfect metches will be seen, es the mismatch target will melt off.

Similarly, variations in buffer composition may be used to elucidate the presence or absence of e mismatch at the detection position. Suitable conditions include, but are not limited to, formamide concentration. Thus, for example, "low" or "permissive" stringency conditions include formamide concentrations of 0 to 10%, while "high" or "stringent" conditions utilize formamide concentrations of ≥40%. Low stringency conditions include NaCl concentretions of ≥ 1 M, and high stringency conditions include concentretions of ≤ 0.3 M. Furthermore, low stringency conditions include $MgCl_2$ concentrations of ≥ 10 mM, moderate stringency as 1-10 mM, and high stringency conditions include concentrations of ≤ 1 mM.

In this embodiment, as for temperature, a plurality of readout probes may be used, with different bases in the readout position (and optionally different labels). Running the asseys under the permissive conditions and repeating under stringent conditions will ellow the elucidation of the base at the detection position.

In one embodiment, the probes used as readout probes are "Molecular Beacon" probes es are generally described in Whitcombe et el., Nature Biotechnology 17:804 (1999), hereby incorporated by reference. As is known in the art, Molecular Beacon probes form "hairpin" type structures, with a fluorescent label on one end and a quencher on the other. In the absence of the target sequence, the ends of the hairpin hybridize, causing quenching of the label. In the presence of a target sequence, the hairpin structure is lost in favor of target sequence binding, resulting in a loss of quenching and thus an increase in signal.

In one embodiment, the Molecular Beecon probes can be the capture probes as outlined herein for readout probes. For example, different beads comprising labeled Molecular Beacon probes (and different bases at the readout position) ere made optionelly they comprise different labels. Alternatively, since Molecular Beacon probes can have spectrally resolvable signals, all four probes (if

e set of four different bases with is used) differently labelled ere ettached to e single beed.

EXTENSION GENOTYPING

In this embodiment, eny number of techniques ere used to edd a nucleotide to the readout position of e probe hybridized to the target sequence adjacent to the detection position. By relying on enzymetic specificity, preferentially e perfectly complementary base is edded. Ali of these methods rely on the enzymatic incorporetion of nucleotides et the detection position. This may be done using chain terminating dNTPs, such that only a single base is incorporated (e.g. singie base extension methods), or under conditions that only a singie type of nucleotide is added foilowed by identification of the edded nucleotide (extension and pyrosequencing techniques).

Single Base Extension

In e preferred embodiment, single base extension (SBE; sometimes referred to es "minisequencing") is used to determine the identity of the base et the detection position. SBE is as described ebove, and utilizes en extension primer that hybridizes to the target nucleic ecid immediately edjacent to the detection position. A polymerase (generally a DNA polymerase) is used to extend the 3' end of the primer with a nucleotide enalog labeled e detection label es described herein. Based on the fidelity of the enzyme, a nucleotide is only incorporated into the readout position of the growing nucleic ecid strand if it is perfectly complementary to the base in the target strand et the detection position. The nucleotide may be derivatized such that no further extensions can occur, so only a single nucleotide is edded. Once the labeled nucleotide is edded, detection of the iabel proceeds es outlined herein.

The reaction is initiated by introducing the assay complex comprising the target sequence (i.e. the array) to a solution comprising e first nucleotide. in general, the nucleotides comprise a detectabie label, which may be either e primary or a secondary label. In addition, the nucieotides mey be nucleotide analogs, depending on the configuration of the system. For example, if the dNTPs are edded in sequential reactions, such that only e single type of dNTP can be added, the nucleotides need not be chain terminating. In eddition, in this embodiment, the dNTPs mey ell comprise the same type of labei.

Alternatively, if the reaction comprises more than one dNTP, the dNTPs shouid be chain termineting, that is, they have a blocking or protecting group at the 3' position such that no further dNTPs may be added by the enzyme. As will be appreciated by those in the art, any number of nucleotide analogs may be used, es long as a polymerase enzyme will still incorporate the nucleotide et the reedout position. Preferred embodiments utiiize dideoxy-triphosphate nucieotides (ddNTPs) end halogenated dNTPs. Generally, e set of nucleotides comprising ddATP, ddCTP, ddGTP and ddTTP is used, each with a different detectable label, eithough es outlined herein, this may not be required. Alternative preferred embodiments use acyclo nucleotides (NEN). These chein terminating nucleotide analogs ere particularly good substrates for Deep vent (exo') and thermosequenase.

In eddition, as will be appreciated by those in the art, the single base extension reactions of the present invention ellow the precise incorporation of modified beses into e growing nucleic ecid strand. Thus, any number of modified nucleotides mey be incorporated for eny number of reasons, including probing structure-function relationships (e.g. DNA:DNA or DNA:protein interactions), cleaving the nucleic ecid, crosslinking the nucleic ecid, incorporate mismetches, etc.

As will be eppreciated by those in the art, the configuration of the genotyping SBE system can take on several forms.

Solution phese essey

As for the OLA reaction described below, the reaction may be done in solution, and then the newly synthesized strands, with the base-specific detectable labels, can be detected. For example, they can be directly hybridized to capture probes that ere complementary to tha extension primers, end the presence of the lebel is then detected. This is schematically depicted in Figure 9C. As will be appreciated by those in the art, a preferred embodiment utilizes four different detecteble labels, i.e. one for each base, such that upon hybridization to the capture probe on the erray, the identification of the base cen be done isothermally. Thus, Figure 9C depicts the readout position **35** es not neccessarily hybridizing to the capture probe.

In a preferred embodiment, edapter sequences cen be used in a solution format. In this embodiment, a single label can be used with a set of four separate primer extension reactions. In this embodiment, the extension reaction is done in solution; each reaction comprises e different dNTP with the label or labeled ddNTP when chain termination is desired. For eech locus genotyped, a set of four different extension primers are used, each with a portion that will hybridize to the target sequence, a different readout base end eech with e different adapter sequence of 15-40 bases, es is more fully outlined below. After the primer extension reaction is complete, the four separate reections era pooled end hybridized to an array comprising complementary probes to the adepter sequences. A genotype is derived by comparing the probe intensities of the four different hybridized adapter sequences corresponding to e give locus.

In eddition, since unextended primers do not comprise labels, the unextended primers need not be removed. However, they may be, if desired, es outlined below; for example, if a large excess of primers are used, there mey not be sufficient signal from the extended primers competing for binding to the surface.

Alternatively, one of skill in the ert could use e single lebel end temperature to determine the identity of the base; thet is, the readout position of the extension primer hybridizes to e position on the capture probe. However, since the three mismetches will have lower Tms than the perfect match, the use of temperature could elucidete the identity of the detection position bese.

Solid phase assay

Alternatively, the reaction may be done on a surface by capturing the target sequence and then running the SBE reection, in a sandwich type format schematically depicted in Figure 9A. In this embodiment, the cepture probe hybridizes to e first domain of the target sequence (which cen be endogeneous or an exogeneous adapter sequenca edded during an amplification reaction), and the extension primer hybridizes to a second target domain immediately adjacent to the detection position. The addition of the enzyme and the required NTPs results in the addition of the interrogation base. In this embodiment, each NTP must have a unique label. Alternatively, each NTP reaction may be done sequentially on a different array. As is known by one of skill in the art, ddNTP and dNTP are the preferred substrates when DNA polymerase is the added enzyme; NTP is the preferred substrate when RNA polymerase is the added enzyme.

Furthermore, as is more fully outlined below and depicted in Figure 9D, capture extender probes cen be used to attach the target sequence to the bead. In this embodiment, the hybridization complex comprises the capture probe, the target sequence end the adapter sequence.

Similarly, the capture probe itself can be used es the extension probe, with its terminus being directly adjacent to the detection position. This is schematicelly depicted in Figure 9B. Upon the addition of the target sequence and the SBE reagents, the modified primer is formed comprising a detectable label, and then detected. Again, as for the solution based reaction, each NTP must have e unique label, the reactions must proceed sequentially, or different arreys must be used. Again, as is known by one of skill in the ert, ddNTP and dNTP are the preferred substrates when DNA polymerase is the added enzyme; NTP is the preferred substrate when RNA polymerase is the added enzyme.

In addition, as outlined herein, the target sequence may be directly ettached to the array; the extension primer hybridizes to it and the reaction proceeds.

Variations on this are shown in Figures 9E end 9F, where the the capture probe and the extension probe adjacently hybridize to the target sequence. Either before or after extension of the extension probe, a ligation step may be used to attach the capture and extension probes together for stability. These are further described below as combination assays.

In addition, Figure 9G depicts the SBE solution reaction followed by hybridization of the product of the reection to the bead array to capture an adapter sequence.

As will be appreciated by those in the art, the determination of the base at the detection position can proceed in several ways. In a preferred embodiment, the reaction is run with all four nucleotides (assuming all four nucleotides are required), each with a different label, as is generally outlined herein. Alternatively, e single label is used, by using four reactions: this may be done either by using a single

substrata and saquantial raactions, or by using four arrays. For axemple, dATP can ba added to the assay complax, and tha generation of a signal evaluatad; tha dATP can be removad and dTTP added, atc. Altamatively, four arrays cen be used; tha first is reacted with dATP, the second with dTTP, etc., and the presance or absence of e signal evaluated. Alternatively, tha raaction includes chain terminating nucleotides such es ddNTPs or acyclo-NTPS.

Alternatively, ratiomatric analysis can be done; for axampla, two labels, "A" and "B", on two substrates (e.g. two arrays) can be done. In this embodiment, two sets of primer extension reactions ara parformed, aach on two arrays, with aach reection containing a complata set of four chain termineting NTPs. Tha first raaction contains two "A" labelad nuclaotides and two "B" labelad nucleotides (for example, A end C may be "A" labalad, and G and T may be "B" labeled). Tha second reaction also contains the two labels, but switched; for axampla, A and G are "A" labalad and T and C are "B" labelad. This reaction composition ellows a biallelic markar to be ratiometrically scored; that is, the intensity of the two labals in two differant "color" channels on a single substrate is compared, using deta from a set of two hybridized arrays. For instance, if tha merker is A/G, then tha first reaction on the first array is used to calculate a ratiometric genotyping score; if the markar is A/C, then the second reaction on the second array is used for the calculation; if the marker is G/T, then the second array is used, etc. This concept can be applied to all possible biallelic marker combinations. "Scoring" e genotype using a singla fiber ratiometric score allows a much mora robust genotyping than scoring a genotypa using a comparison of absolute or normalized intensities between two different arrays.

Removal of unextended primars

In a preferred embodiment, for both SBE as wall as a number of other reactions outlined hereln, it is desirabla to remove tha unextended or unreactad primers from the assay mixture, end particularly from the array, as unextended primers will compete with tha extended (labeled) primers in binding to capture probes, thereby diminishing the signal. The concentration of the unextended primers ralativa to the extended primer may ba ralatively high, since a large excess of primer is usually required to generate efficient primer annealing. Accordingly, a number of different tachniquas may be used to facilitate the removal of unaxtended primers. As outlined above, these genarally include methods based on removal of unreacted primers by binding to a solid support, protecting the reacted primers and degrading the unextandad ones, and separating the unreacted and reacted primars.

Protection and degradation

In this embodiment, tha ddTNPs or dNTPs that are added during the reaction confar protection from degradation (whether chemical or enzymatic). Thus, after the assay, the dagradation components ere added, and unreacted primers ere degraded, leaving only tha reactad primars. Labeled protecting groups are particularly praferrad; for example, 3'-substituted-2'-dNTPs can contain anthranylic derivatives that ere fluorescent (with alkali or enzymatic treatmant for removal of the protecting group).

In a preferred embodiment, the secondary label is a nuclease inhibitor, such as thiol NTPs. In this embodiment, the chain-terminating NTPs are chosen to render extended primers resistant to nucleases, such as 3'-exonucleases. Addition of an exonuclease will digest the non-extended primers leaving only the extended primers to bind to the capture probes on the array. This may also be done with OLA, wherein the ligated probe will be protected but the unprotected ligation probe will be digested.

In this embodiment, suitable 3'-exonucleases include, but are not limited to, exo I, exo III, exo VII, and 3'-5' exophosphodiestereses.

Alternatively, an 3' exonuclease may be added to a mixture of 3' labeled biotin/streptavidin; only the unreacted oligonucleotides will be degraded. Following exonuclease treatment, the exonuclease and the streptavidin can be degraded using a protease such as proteinase K. The surviving nucleic acids (i.e. those that were biotinylated) are then hybridized to the array.

Separation systems

The use of secondary label systems (and even some primary label systems) can be used to separate unreacted and reacted probes; for example, the addition of streptavidin to a nucleic acid greatly increases its size, as well as changes its physical properties, to allow more efficient separation techniques. For example, the mixtures can be size fractioned by exclusion chromatography, affinity chromatography, filtration or differential precipitation.

Non-terminated extension

In a preferred embodiment, methods of adding a single base are used that do not rely on chain termination. That is, similar to SBE, enzymatic reactions that utilize dNTPs and polymerases can be used; however, rather than use chain terminating dNTPs, regular dNTPs are used. This method relies on a time-resolved basis of detection; only one type of base is added during the reaction. Thus, for example, four different reactions each containing one of the dNTPs can be done; this is generally accomplished by using four different substrates, although as will be appreciated by those in the art, not all four reactions need occur to identify the nucleotide at a detection position. In this embodiment, the signals from single additions can be compared to those from multiple additions; that is, the addition of a single ATP can be distinguished on the basis of signal intensity from the addition of two or three ATPs. These reactions are accomplished as outlined above for SBE, using extension primers and polymerases; again, one label or four different labels can be used, although as outlined herein, the different NTPs must be added sequentially.

A preferred method of extension in this embodiment is pyrosequencing.

Pyrosequencing

Pyrosequencing is an extension end sequencing method that cen be used to edd one or more nucleotides to the detection position(s); it is very similar to SBE except thet chain terminating NTPs need not be used (elthough they may be). Pyrosequencing relies on the detection of e reection product, PPI, produced during the eddition of en NTP to a growing oligonucleotide chain, rather than on e label ettached to the nucleotide. One molecule of PPI is produced per dNTP edded to the extension primer. Thet is, by running sequential reactions with eech of the nucleotides, end monitoring the reection products, the identity of the edded base is determined.

The release of pyrophosphete (PPi) during the DNA polymerase reaction can be quantitatively meesured by meny different methods end a number of enzymetic methods have been described; see Reeves et el., Anel. Biochem. 28:282 (1969); Guillory et el., Anel. Biochem. 39:170 (1971); Johnson et al., Anal. Biochem. 15:273 (1968); Cook et el., Anal. Biochem. 91:557 (1978); Drake et el., Anal. Biochem. 94:117 (1979); WO93/23564; WO 98/28440; WO98/13523; Nyren et al., Anal. Biochem. 151:504 (1985); ell of which are incorporated by reference. The letter method allows continuous monitoring of PPi and has been termed ELIDA (Enzymebc Luminometric Inorganic Pyrophosphate Detection Assay). A preferred embodiment utilizes eny method whlch can result in the generation of en optical signal, with preferred embodiments utilizing the generetion of e chemlluminescent or fluorescent signal.

A preferred method monitors the creation of PPi by the conversion of PPI to ATP by the enzyme sulfurylase, and the subsequent production of visible light by firefly luciferase (see Ronaghi et el., Science 281:363 (1998), incorporated by reference). In this method, the four deoxynucleotides (dATP, dGTP, dCTP end dTTP; collectively dNTPs) are added stepwise to e partial duplex comprising a sequencing primer hybridized to a single strended DNA template and incubated with DNA polymerase, ATP sulfurylase, luciferase, end optionelly a nucleotide-degrading enzyme such es epyrase. A dNTP is only incorporeted into the growing DNA strand if it is complementary to the base in the template strand. The synthesis of DNA is accompanied by the release of PPi equal in molarity to the incorporeted dNTP. The PPi is converted to ATP end the light genereted by the luciferase is directly proportional to the amount of ATP. In some cases the unincorporated dNTPs end the produced ATP are degraded between each cycle by the nucleotide degrading enzyme.

Accordingly, a preferred embodiment of the methods of the invention is es follows. A substrate comprising microspheres containing the target sequences end extension primers, forming hybridizetion complexes, is dipped or contacted with e reection volume (chamber or well) comprising a single type of dNTP, en extension enzyme, end the reegents end enzymes necessary to detect PPi. If the dNTP is complementery to the base of the target portion of the target sequence edjacent to the extension primer, the dNTP is edded, releesing PPI and generating detectable light, which is detected as generally described in U.S.S.N.s 09/151,877 and 09/189,543, and PCT US98/09163, all of which are hereby incorporeted by reference. If the dNTP is not complementary, no detectable signal results.

The substrate is than contacted with a second reaction volume (chamber) comprising e different dNTP and tha edditional components of tha assay. This process is rapaeted if the identity of a basa at a sacond datection position is dasirabla.

In a preferrad embodiment, washing staps, i.a. tha usa of washing chambers, may be dona in betwean tha dNTP raaction chambers, es raquirad. Thase washing chambers may optionally comprise e nucleotide-degrading enzyma, to remova any unreactad dNTP and decreasing the background signal, as is described in WO 98/28440, incorporated herein by referenca.

As will be eppraciated by thosa in tha art, the system can be configured in a variaty of ways, including both a linaer progression or a circular ona; for example, four arrays may ba used that each can dip into one of four reaction chambars arrayed in a circular pattarn. Each cycla of sequencing and reading is followed by e 90 dagree rotation, so thet aach substrata then dips into the naxt raaction wall.

In a preferred embodiment, one or more internal control saquencas are used. That is, at least ona microsphere in the array comprises a known sequence that can be used to vanify that tha reactions ere proceeding correctly. In a preferrad embodiment, at least four control sequences are usad, eech of which has a differant nucleotide at eech position: the first control sequence will have an adenosina at position 1, tha second will have e cytosine, tha third e guanosine, and tha fourth a thymidine, thus ensuring that at least one control sequence is "lighting up" at each step to serve es en internal control.

As for simple axtension and SBE, the pyrosequencing systems mey be configured in e variety of ways; for example, the target sequence mey be attached to the bead in a variety of ways, including direct attachment of the target sequence; the use of a capture probe with e separete extension probe; tha usa of a capture extender probe, a capture proba and a separata extension probe; the usa of edapter sequences in the target sequenca with cepture and extension probes; and the use of e capture probe that also serves es the extension probe.

Ona additional benafit of pyrosequencing for genotyping purposes is that since tha reection does not rely on tha incorporation of labels into a growing chain, the unreacted extension primers need not be removed.

Allelic PCR

In e preferred embodiment, tha method used to detect tha base at the detection position is allelic PCR, referred to herein as "aPCR". As described in Newton at el., Nucl. Acid Ras. 17:2503 (1989), hereby expressly incoporated by reference, ellelic PCR ellows singla besa discriminetion based on the fact that the PCR raaction does not proceed well if the terminal 3'-nucleotide is mismatched, assuming the DNA polymerase being used lacks a 3'-axonucleasa proofreading activity. Accordingly, the identification of the base proceeds by using allelic PCR primers (sometimes refarred to herein as

aPCR primers) that hava readout positions et their 3' ends. Thus the target sequence comprises a first domeln comprising et its 5' end a detection position.

In generel, ePCR may be briefly described es follows. A double stranded target nuclelc acid is denatured, generally by raising the temperature, and then cooled In the prasence of an excess of a aPCR primer, which then hybridizes to the first target strand. If the readout position of the aPCR primer basepairs correctly with tha detection position of the target sequence, a DNA polymerase (egaln, that lacks 3'-exonuclease activity) then acts to extend the primer with dNTPs, resulting in the synthasis of a naw strand forming a hybridization complex. Tha sample is then heated again, to disassociate the hybridization complax, and the process is repeated. By using e second PCR primer for tha complementary target strand, rapid and exponential amplification occurs. Thus aPCR steps are denaturation, anneeling and extension. The particulars of aPCR ere wall known, and include the use of a thermostable polymerese such es Taq I polymerese and thermel cycling.

Accordingly, the aPCR reaction requires et least one ePCR primer, a polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

Furthermore, the ePCR reaction may be run as a competition assay of sorts. For example, for biallelic SNPs, a first aPCR primer comprising a first base at the readout position end a first label, and a second aPCR primer comprising a different base at the readout position and a second label, may be used. The PCR primer for the other strand is the same. The examinetion of the ratio of the two colors can serve to identify the base et the detection position.

In general, as is more fully outlined below, the capture probes on the beads of the array are designed to be substantially complementary to the extended part of the primer; thet is, unextended primers will not bind to the capture probes.

LIGATION TECHNIQUES FOR GENOTYPING

In this embodiment, the readout of tha base at the detection position proceeds using a ligase. In this embodiment, it is the specificity of tha ligase which is the besis of the genotyping; that is, ligases generally require that the 5' and 3' ends of the ligetion probes heve perfect complamentarity to the target for ligation to occur. Thus, in a preferred embodiment, the identity of the bese et the detection position proceeds utilizing OLA as described ebove, es is generally depicted in Figure 10. Tha method can be run at least two different ways; in a first embodiment, only one strand of a target sequence is used es a templete for ligation; eltemetively, both strends may be used; the latter is generally referred to as Ligation Chain Reaction or LCR.

This method is based on the fact that two probes can be preferentially ligated together, if they are

hybridized to e terget strand end if perfect complementarity exists at the two bases being ligated together. Thus, in this embodiment, the target sequence comprises e contiguous first target domein comprising the detection position end e second terget domeln edjecent to the detection position. That is, the detection position is "between" the rest of the first terget domaln and the second target domain. A first ligetion probe is hybridized to the first target domain and a second ligetion probe is hybridized to the second target domain. If the first ligation probe hes a base perfectly complementary to the detection position base, and the adjacent bese on the second probe hes perfect complementarity to its position, e ligetion structure is formed such thet the two probes can be ligated together to form a ligeted probe. If this complementarity does not exist, no ligation structure is formed end the probes are not ligated together to an appreciable degree. This may be done using heat cycling, to allow the ligated probe to be denatured off the target sequence such thet it may serve es a template for further reactions. In addition, es is more fully outlined below, this method may elso be done using ligetion probes that are separated by one or more nucleotides, if dNTPs and a polymerase are added (this is sometimes referred to es "Genetic Bit" enelysis).

In a preferred embodiment, LCR is done for two strands of e double-strended target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strend of the target, and a separate set (i.e. third end fourth primer probe nucleic ecids) for the other strend of the target. In a preferred embodiment, the first end third probes will hybridize, end the second and fourth probes will hybridize, such that amplification can occur. That is, when the first and second probes have been attached, the ligated probe can now be used as a templete, in eddition to the second target sequence, for the attachment of the third end fourth probes. Similarly, the ligated third and fourth probes will serve as a template for the attachment of the first and second probes, in addition to the first terget strand. In this way, en exponential, rather than just a lineer, amplification can occur.

As will be appreciated by those in the ert, the ligation product can be detected in a variety of ways. Preferably, detection is eccomplished by removing the unligated labeled probe from the reaction before applicetion to e capture probe. In one embodiment, the unligeted probes are removed by digesting 3' non-protected oligonucleotides with e 3' exonuclease, such es, exonuclease I. The ligation products ere protected from exo I digestion by including, for example, the use of e number of sequential phosphorothioate residues at their 3' terminus (for example et leest four), thereby, rendering them resistant to exonuclease digestion. The unligeted detection oligonucleotides ere not protected and are digested.

As for most or ell of the methods described herein, the essay can take on a solution-based form or a solid-phase form.

Solution based OLA

In a preferred embodimant, as shown in Figura 10A, tha ligation reaction is run in solution. In this ambodiment, only one of the primers carries a detectable label, e.g. the first ligetion probe, end the capture proba on the bead is substantially complementary to the other probe, e.g. the second ligation proba. In this way, unextended labeled ligation primers will not intarfara with the assay. This substantielly reducas or eliminates false signel generated by the optically-labeled 3' primers.

In addition, a solution-based OLA assay that utilizes adaptar sequences may be done. In this embodiment, rather than heve tha target sequence comprise the edepter sequences, one of the ligation probes comprises the adapter sequence. This facilitates the creation of "universal arrays". For axample, es depicted in Figure 10E, tha first ligation probe has an adapter sequence that is used to attach the ligated probe to the array.

Again, as outlined above for SBE, unreacted ligetion primers may be removad from the mixture as naeded. For example, the first ligation probe may comprise the label (either a primary or secondary label) and the second may be blocked at its 3' end with an exonucleasa blocking moiety; after ligation and the introduction of the nuclease, the labaled ligation proba will be digested, leaving the ligation product and the second probe; however, since the second proba is unlabeled, it is effectivaly silant in the assey. Similarly, the second probe mey comprise e binding partnar used to pull out the ligeted probes, leaving unligated labeled ligation probas behind. The binding pair is then disassociated and edded to the errey.

Solid phase based OLA

Altarnatively, tha targat nucleic ecid is immobilized on a solid-phase surface. The OLA essay is performed and unligated oligonucleotides are removed by washing under eppropriate stringency to remove unligated oligonucleotides and thus tha label. For exampla, as depicted in Figure 10B, the capture probe can comprisa one of the ligetion probes. Similarly, Figures 10C and 10D depict alternative attachments.

Again, as outlined ebove, the detection of the OLA reaction can also occur directly, in the case where one or both of the primers comprises at least one detectable label, or indirectly, using sandwich assays, through the use of edditional probes; that is, tha ligated probes can serve es target sequencas, end detection may utilize amplification probes, capture probas, captura extender probas, label probes, and label extender probes, etc.

Solid Phase Oligonucleotida Ligation Assay (SPOLA)

In a preferred embodiment, a novel method of OLA is used, tarmed herein "solid phasa oligonucleotida assay", or "SPOLA". In this embodimant, the ligation probas ara both attached to the same site on the surface of the erray (e.g. when microsphere arrays are used, to the same bead), one at its 5' end (the "upstream probe") and one at its 3' end (the "downstream probe"), es is generally

depicted in Figure 11. This may be done as is will be appreciated by those in the art. At least one of the probes is attached via a cleaveble linker, that upon cleavage, forms a reactive or detectable (fluorophore) moiety. If ligation occurs, the reactive moiety remains associated with the surface; but if no ligation occurs, due to a mismatch, the reactive moiety is free in solution to diffuse away from the surface of the array. The reactive moiety is then used to add a detectable label.

Generally, as will be appreciated by those in the art, cleavage of the cleavable linker should result in asymmetrical products; i.e. one of the "ends" should be reactive, and the other should not, with the configuration of the system such that the reactive moiety remains associated with the surface if ligation occurred. Thus, for example, amino acids or succinate esters can be cleaved either enzymatically (via peptidases (aminopeptidase and carboxypeptidase) or proteases) or chemically (acid/base hydrolysis) to produce an amine and a carboxyl group. One of these groups can then be used to add a detectable label, as will be appreciated by those in the art and discussed herein.

Padlock probe ligation

In a preferred embodiment, the ligation probes are specialized probes called "padlock probes". Nilsson et al, 1994, Science 265:2085, hereby incorporated by reference. These probes have a first ligation domain that is identical to a first ligation probe, in that it hybridizes to a first target sequence domain, and a second ligation domain, identical to the second ligation probe, that hybridizes to an adjacent target sequence domain. Again, as for OLA, the detection position can be either at the 3' end of the first ligation domain or at the 5' end of the second ligation domain. However, the two ligation domains are connected by a linker, frequently nucleic acid. The configuration of the system is such that upon ligation of the first and second ligation domains of the padlock probe, the probe forms a circular probe, and forms a complex with the target sequence wherein the target sequence is "inserted" into the loop of the circle.

In this embodiment, the unligated probes may be removed through degradation (for example, through a nuclease), as there are no "free ends" in the ligated probe.

CLEAVAGE TECHNIQUES FOR GENOTYPING

In a preferred embodiment, the specificity for genotyping is provided by a cleavage enzyme. There are a variety of enzymes known to cleave at specific sites, either based on sequence specificity, such as restriction endonucleases, or using structural specificity, such as is done through the use of invasive cleavage technology.

ENDONUCLEASE TECHNIQUES

In a preferred embodiment, enzymes that rely on sequence specificity are used. In general, these systems rely on the cleavage of double stranded sequence containing a specific sequence recognized by a nuclease, preferably an endonuclease including resolvases.

Thesa systams may work in a variaty of ways, as is genarally depictad in Figura 12. In one embodimant (Figure 12A), a labelad raadout proba (genarally ettachad to e baad of tha erray) is usad; tha binding of tha targat saquanca forms a doubla strended sequenca that a rastriction endonuclaesa can than recogniza and claeva, if tha corract sequance is prasant. An anzyma resulting in "sticky ends" is shown in Figure 12A. The cleavaga results in the loss of the lebel, and thus e loss of signel.

Altarnativaly, as will ba appreciated by thosa in the ert, a labailed targat sequanca may be used as well; for example, a labellad primar may be usad in tha PCR emplification of tha target, such that tha label is incorporatad in such a mannar as to be cleaved off by tha anzyma.

Alternativaly, the raadout probe (or, again, the target sequance) may comprise both a fluorascent label and e quencher, as is known in tha art and depicted in Figure 12B. In this embodiment, the label and tha quencher are attached to diffarant nucleosidas, yet are close enough that tha quencher molecule rasults in littla or no signal being presant. Upon the introduction of the anzyma, the quencher is claaved off, leaving tha labal, end allowing signalling by tha label.

In addition, as will ba appreciated by those in the art, thesa systems can ba both solution-based assays or solid-phase asseys, es outlined herain.

Furthermore, thera ere some systams that do not require cleavage for detection; for example, some nucleic ecid binding protains will bind to specific sequances and can thus serve as a sacondery label. For example, some transcription factors will bind in e highly saquence dependent mannar, and can distinguish batwaen two SNPs. Having bound to the hybridization complex, a detactable binding partnar can be eddad for detaction. In addition, mismatch binding proteins based on mutated transcription factors can be used.

In eddition, as will be appreciated by thosa in the ert, this type of approach works with other cleavage methods as well, for example tha use of invasiva cleavage methods, es outlined below.

Invasive cleevega

In a preferred embodiment, the datermination of tha identity of the basa at the detection position of the target saquence proceeds using invasive cleavaga tachnology. As outlined abova for emplification, invasive cleavaga techniquas rely on tha use of structure-specific nucleasas, whara the structure can be formad as a rasult of tha prasance or absence of a mismatch. Generally, invasive claavage technology may ba described as follows. A target nucleic acid is recognized by two distinct probes. A first probe, genarally referred to herein as an "invader" probe, is substantielly complementary to a first portion of the targat nucleic acid. A second probe, genarally rafarred to herein as a "signal proba", is partially complamentary to the target nucleic acid; tha 3' end of tha signal oligonuclaotida is substantially complemantary to tha target sequance whila tha 5' end is non-complementary and

preferebly forms e single-stranded "tail" or "erm". The non-complementary end of the second probe preferebly comprises e "generic" or "unique" sequence, frequently referred to herein es e "detection sequence", that is used to indicate the presence or ebsence of the target nucleic ecid, es described below. The detection sequence of the eecond probe preferably comprises et leest one detectable label. Altemetive methods have the detection sequence functioning es e target sequence for e capture probe, and thus rely on sandwich configurations using label probes.

Hybridization of the first end second oligonucleotides neer or adjacent to one enother on the target nucleic ecid forms e number of structures. In e preferred embodiment, a forked cleevage structure, as shown in Figure 13, forms end is e substrate of e nucleese which cleeves the detection sequence from the signal oligonucleotide. The site of cleevage is controlled by the distance or overlap between the 3' end of the invader oligonucleotide and the downstream fork of the signal oligonucleotide. Therefore, neither oligonucleotide is subject to cleavege when misaligned or when unettached to target nucleic acid.

As ebove, the invasive cleavege essay is preferably performed on an errey formet. In e preferred embodiment, the signal probe has e detectable label, ettached 5' from the site of nuclease cleavage (e.g. within the detection sequence) end a capture tag, es described herein for removal of the unreacted products (e.g. biotin or other hapten) 3' from the site of nucleese cleevege. After the essay is cerried out, the uncleeved probe end the 3' portion of the cleeved signal probe (e.g. the the detection sequence) mey be extracted, for exemple, by binding to streptavidin beeds or by crosslinking through the cepture tag to produce eggregates or by antibody to en ettached hepten. By "capture teg" herein is e meant one of a pair of binding partners as described above, such es entigen/entibody pairs, digoxygenenin, dinitrophenol, etc.

The cleaved 5' region, e.g. the detection sequence, of the signal probe, comprises e lebel end is detected end optionally quantitated. In one embodiment, the cleeved 5' region is hybridized to a probe on en array (capture probe) and opticelly detected (Figure 13). As described below, many different signel probes cen be enelyzed in parellel by hybridization to their complementary probes in an array. In e preferred embodiment es depicted in Figure 13, combination techniques ere used to obtain higher specificity and reduce the detection of contamineting uncleeved signel probe or incorrectly cleeved product, an enzymatic recognition step is introduced in the erray capture procedure. For example, es more fully outlined below, the cleaved signal probe binds to e cepture probe to produce e double-strended nucleic ecid in the erray. In this embodiment, the 3' end of the cleaved signal probe is edjacent to the 5' end of one strand of the capture probe, thereby, forming e substrete for DNA ligese (Broude et el. 1991. PNAS 91: 3072-3076). Only correctly cleaved product is ligeted to the capture probe. Other incorrectly hybridized end non-cleeved signel probes are removed, for example, by heet denaturation, high stringency washes, end other methods thet disrupt base pairing.

Accordingly, the present invention provides methods of determining the identity of a base at the detection position of a target sequence. In this embodiment, the target sequence comprises, 5' to 3', a first target domain comprising an overlap domain comprising at least a nucleotide in the detection position, and a second target domain contiguous with the detection position. A first probe (the "invader probe") is hybridized to the first target domain of the target sequence. A second probe (the "signal probe"), comprising a first portion that hybridizes to the second target domain of the target sequence and a second portion that does not hybridize to the target sequence, is hybridized to the second target domain. If the second probe comprises a base that is perfectly complementary to the detection position a cleavage structure is formed. The addition of a cleavage enzyme, such as is described in U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,029; 5,541,311 and 5,843,669, all of which are expressly incorporated by reference, results in the cleavage of the detection sequence from the signalling probe. This then can be used as a target sequence in an assay complex.

In addition, as for a variety of the techniques outlined herein, unreacted probes (i.e. signalling probes, in the case of invasive cleavage), may be removed using any number of techniques. For example, the use of a binding partner (**70** in Figure 13C) coupled to a solid support comprising the other member of the binding pair can be done. Similarly, after cleavage of the primary signal probe, the newly created cleavage products can be selectively labeled at the 3' or 5' ends using enzymatic or chemical methods.

Again, as outlined above, the detection of the invasive cleavage reaction can occur directly, in the case where the detection sequence comprises at least one label, or indirectly, using sandwich assays, through the use of additional probes; that is, the detection sequences can serve as target sequences, and detection may utilize amplification probes, capture probes, capture extender probes, label probes, and label extender probes, etc.

In addition, as for most of the techniques outlined herein, these techniques may be done for the two strands of a double-stranded target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strand of the target, and a separate set for the other strand of the target.

Thus, the invasive cleavage reaction requires, in no particular order, an invader probe, a signalling probe, and a cleavage enzyme.

As for other methods outlined herein, the invasiva cleavage reaction may be done as a solution based assay or a solid phase assay.

Solution-based invasive cleavage

The invasive cleavage reaction may be done in solution, followed by addition of one of the

components to an array, with optionel (but preferable) removal of unreected probee. For exemple, es depicted in Figure 13C, the reaction is carried out in solution, using e ceptura tag (i.e. e member of e binding partner peir) that is separeted from the label on the detection sequence with the cleavage site. After cleevage (dependent on the base at the detection position), the signelling probe is cleeved. The ceptura tag is used to remove the uncleaved probes (for example, using megnatic perticlas comprising the other member of the binding pair), end the remaining solution is added to tha erray. Figura 13C depicts the direct ettachmant of the detection sequence to the capture proba. In this embodimant, the detaction sequence can effectivaly act es an edapter sequence. In alternete embodimants, ae dapicted in Figure 13D, tha detection sequance is unlabelled end an edditionel label probe is used; as outlined below, this cen be ligated to the hybridizetion complex.

Solid-phase besed essays

The invasive cleevage reection can also be done es a solid-phese assey. As dapicted in Figura 13A, the target sequenca can be attached to the array using a captura probe (in addition, elthough not shown, tha target sequence may be directly attached to the array). In a preferred embodiment, the signalling probe comprises both a fluorophore label (attached to the portion of the signalling probe that hybridizes to the target) and e quencher (generally on the detection sequence), with a cleavage site in between. Thus, in the absence of cleavage, very little signel is seen due to the quenching reaction. After cleavage, howaver, tha detection sequance is removed, elong with the quencher, leaving the unquenched fluorophora. Similarly, the invesive probe may be attached to the array, as depicted in Figure 13B.

In a preferred embodiment, the invasive cleavage reaction is configured to utilize a fluorophore-quencher reaction. A signalling probe comprising both e fluorophore end e quencher is ettached to the bead. The fluorophore is contained on the portion of the signalling probe that hybridizes to the target sequence, and the quencher is contained on a portion of the signalling probe that is on the other side of the cleavage site (termed the "detection sequance" herein). In a praferred ambodiment, it is tha 3' end of the signalling probe that is attached to the beed (although as will be eppreciated by those in the art, the system can be configured in a variety of different ways, including methods that would result in a loss of signal upon cleavaga). Thus, the quenchar molecule is loceted 5' to the cleavage site. Upon assembly of an assay complex, comprising the target sequence, an invader probe, end e signelling probe, and the introduction of the cleavage enzyme, the cleavage of tha complex results in the disassociation of the quencher from the complex, resulting in en increase in fluorescence.

In this embodiment, suitable fluorophore-quencher pairs are as known in the ert. For example, suitable quencher molecules comprise Dabcyl.

COMBINATION TECHNIQUES

It is elso possible to combine two or more of these techniques to do genotyping, quantification,

detection of sequences, etc.

Novel combination of competitive hybridization and extension

In a preferred embodiment, a combination of competitive hybridization and extension, particularly SBE, is used. This may be generally described as follows. In this embodiment, different extension primers comprising different bases at the readout position are used. These are hybridized to a target sequence under stringency conditions that favor perfect matches, and then an extension reaction is done. Basically, the readout probe that has the match at the readout position will be preferentially extended for two reasons; first, the readout probe will hybridize more efficiently to the target (e.g. has a slower off rate), and the extension enzyme will preferentially add a nucleotide to a "hybridized" base. The reactions can then be detected in a number of ways, as outlined herein.

The system can take on a number of configurations, depending on the number of labels used, the use of adapters, whether a solution-based or surface-based assay is done, etc. Several preferred embodiments are shown in Figure 14.

In a preferred embodiment, at least two different readout probes are used, each with a different base at the readout position and each with a unique detectable label that allows the identification of the base at the readout position. As described herein, these detectable labels may be either primary or secondary labels, with primary labels being preferred. As for all the competitive hybridization reactions, a competition for hybridization exists with the reaction conditions being set to favor match over mismatch. When the correct match occurs, the 3' end of the hybridization complex is now double stranded and thus serves as a template for an extension enzyme to add at least one base to the probe, at a position adjacent to the readout position. As will be appreciated by those in the art, for most SNP analysis, the nucleotide next to the detection position will be the same in all the reactions.

In one embodiment, chain terminating nucleotides may be used; alternatively, non-terminating nucleotides may be used and multiple nucleotides may be added, if desired. The latter may be particularly preferred as an amplification step of sorts; if the nucleotides are labelled, the addition of multiple labels can result in signal amplification.

In a preferred embodiment, the nucleotides are analogs that allow separation of reacted and unreacted primers as described herein; for example, this may be done by using a nuclease blocking moiety to protect extended primers and allow preferentially degradation of unextended primers or biotin (or iminobiotin) to preferentially remove the extended primers (this is done in a solution based assay, followed by elution and addition to the array).

As for the other reactions outlined herein, this may be done as a solution based assay, or a solid phase assay. Solution based assays are generally depicted in Figures 14A, 14B and 14C. In a solid

phase reaction, an example of which is depicted in Figure 14D, the capture probe serves as the readout probe; In this embodiment, different positions on the array (e.g. different beeds) comprise different readout probes. Thet la, at least two different capture/readout probes are used, with three and four also possible, depending on the allele. The reaction is run under conditions that favor the formation of perfect match hybridization complexes. In this embodiment, the dNTPs comprise e detectable lebel, preferably e primary label such as a fluorophore. Since the competitive readout probes ere spatielly defined In the array, one fluorescent label can distinguish between the alleles; furthermore, it is the same nucleotide that is being added in the reaction, since it is the position edjacent to the SNP thet is being extended. As for ell the competitive assays, reletive fluorescence intensity distinguishes between the alleles and between homozygosity and heterozygosity. In eddition, multiple extension reactions can be done to amplify the signal.

For both solution and solid phase reactions, edepters may be additionally used. In e preferred embodiment, as shown in Figure 14B for the solution based assay (elthough es will be appreciated by those in the ert, a solid phase reaction mey be done as well), adapters on the 5' ends of the readout probes ere used, with identical adapters used for each allele. Each readout probe has a unique detectable label that ellows the determination of the base et the reedout position. After hybridizetion and extension, the readout probes are added to the arrey; the adapter sequences direct the probes to particular array locations, and the relative intensities of the two labels distinguishes between alleles.

Alternatively, es depicted in Figure 14C for the solution based assey (although as will be eppreciated by those in the art, a solid phese reaction mey be done as well), a different adapter may be used for each readout probe. In this embodiment, e single label may be used, since spatial resolution is used to distinguish the elleles by having a unique adapter attached to each allelic probe. After hybridization and extension, the readout probes are added to the arrey; the unique adepter sequences direct the probes to unique array locations. In this embodiment, it is the relative intensities of two array positions that distinguishes between alleles.

As will be appreciated by those in the ert, eny array may be used in this novel method, including both ordered and random arrays. In a preferred embodiment, the arrays may be made through spotting techniques, photolithographic techniques, printing techniques, or preferably ere bead arrays.

Combination of competitive hybridization end invasive cleavage

In a preferred embodiment, e combination of competitive hybridizetion end invasive cleevage is done. As will be appreciated by those in the ert, this technique is invasive cleevege as described above, with at least two sets of probes comprising different bases in the readout position. By running the reactions under conditions that favor hybridization complexes with perfect matches, different alleles may be distinguished.

In e preferred embodiment, this technique is done on bead errays.

Novel combination of invesive cleavege end ligation

In e preferred embodiment, Invasive cleavage and ligation is done, as is generelly depicted in Figure 15. In this embodiment, the specificity of the invasive cleavage reaction is used to detect the nucleotide in the detection position, end the specificity of the ligase reaction is used to ensure that only cleaved probes give e signal; that is, the ligation reaction confers an extra level of specificity.

The detection sequence, comprising e detecteble label, of the signal probe is cleaved if the correct basepairing is present, es outlined above. The detection sequence then serves as the "target sequence" in a secondary reaction for detection; it is added to e capture probe on a microsphere. The capture probe in this case comprises e first double stranded portion and a second single stranded portion that will hybridize to the detection sequence. Again, preferred embodiments utilize edjacent portions, although dNTPs and a polymerase to fill in the "gep" may also be done. A ligase is then added. As shown in Figure 15A, only if the signal probe hes been cleaved will ligation occur; this results in covalent attachment of the signal probe to the array. This may be detected es outlined herein; preferred embodiments utilize stringency conditions that will discriminate between the ligated and unligated systems.

As will be appreciated by those in the art, this system may take on a number of configurations, including solution based and solid based assays. In a preferred embodiment, as outlined above, the system is configured such that only if cleevage occurs will ligation happen. In a preferred embodiment, this may be done using blocking moieties; the technique can generally be described as follows. An invasive cleavage reaction is done, using a signalling probe that is blocked at the 3' end. Following cleavage, which creates a free 3' terminus, e ligation reaction is done, generally using a template target and a second ligation probe comprising a detectable label. Since the signalling probe has a blocked 3' end, only those probes undergoing cleavage get ligated and labelled.

Alternatively, the orientations may be switched; in this embodiment, a free 5' phosphate is generated and is available for labeling.

Accordingly, in this embodiment, a solution invasive cleavage reaction is done (although as will be appreciated by those in the ert, a support bound invasive cleavage reaction may be done es well).

As will be appreciated by those in the art, eny errey may be used in this novel method, including both ordered (predefined) and rendom erreys. In a preferred embodiment, the arreys may be made through spotting techniques, photolithographic techniques, printing techniques, or preferably are bead arreys.

Combination of invasive cleavage end extension

In a preferred embodiment, a combination of invasive cleavage end extension reactions are done, as generally depicted in Figure 16. The technique can generally be described as follows. An invasive cleavage reaction is done, using a signalling probe that is blocked at the 3' end. Following cleavage, which creates a free 3' terminus, an extension reaction is done (either enzymatically or chemically) to add a detectable label. Since the signalling probe has a blocked 3' end, only those probes undergoing cleavage get labelled.

Alternatively, the orientations may be switched, for example when chemical extension or labeling is done. In this embodiment, a free 5' phosphate is generated and is available for labeling.

In a preferred embodiment, the invasive cleavage reaction is configured as shown in Figure 16B. In this embodiment, the signalling probe is attached to the array at the 5' end (e.g. to the detection sequence) and comprises a blocking moiety at the 3' end. The blocking moiety serves to prevent any alteration (including either enzymatic alteration or chemical alteration) of the 3' end. Suitable blocking moieties include, but are not limited to, chain terminators, alkyl groups, halogens; basically any non-hydroxy moiety.

Upon formation of the assay complex comprising the target sequence, the invader probe, and the signalling probe, and the introduction of the cleavage enzyme, the portion of the signalling probe comprising the blocking moiety is removed. As a result, a free 3' OH group is generated. This can be extended either enzymatically or chemically, to incorporate a detectable label. For example, enzymatic extension may occur. In a preferred embodiment, a non-templated extension occurs, for example, through the use of terminal transferase. Thus, for example, a modified dNTP may be incorporated, wherein the modification comprises the presence of a primary label such as a fluor, or a secondary label such as biotin, followed by the addition of a labeled streptavidin, for example. Similarly, the addition of a template (e.g. a secondary target sequence that will hybridize to the detection sequence attached to the bead) allows the use of any number of reactions as outlined herein, such as simple extension, SBE, pyrosequencing, OLA, etc. Again, this generally (but not always) utilizes the incorporation of a label into the growing strand.

Alternatively, as will be appreciated by those in the art, chemical labelling or extension methods may be used to label the 3' OH group.

As for all the combination methods, there are several advantages to this method. First of all, the absence of any label on the surface prior to cleavage allows a high signal-to-noise ratio. Additionally, the signalling probe need not contain any labels, thus making synthesis easier. Furthermore, because the target-specific portion of the signalling probe is removed during the assay, the remaining detection sequence can be any sequence. This allows the use of a common sequence for all beads; even if

different reactions ere carried out in parallel on the array, the post-cleavage detection can be identical for all assays, thus requiring only one set of reegents. As will be appreciated by those in the art, it is also possible to have different detection eequences if required. In addition, since the label is attached post-cleavage, there is a great deal of flexibility in the type of label thet may be incorporated. This can leed to significant signal emplification; for example, the use of highly labeled streptavidin bound to a blotin on the detection sequence can give an increesed signal per detection sequence. Similarly, the use of enzyme labels such as alkaline phosphatase or horseradish peroxidase allow signal amplification as well.

A further adventage is the two-fold specificity that is built into the essay. By requiring spacificity at the cleavage etep, followed by specificity at the extension step, increased signal-to-noisa ratios are seen.

As will be appreciated by those in the art, while generally dascribed as a solid phase assey, this reaction may also be done in solution; this is similar to the solution-besed SBE reactions, wherein the detection sequence serves as the extension primer. This assay also may be performed with an extension primer/adaptor oligonucleotide es described for solution-based SBE assays. It should be noted that the arrays used to detect the invasive cleavage/extension reactions may be of any type, including, but not limited to, spotted and printed arrays, photolithographic arrays, and bead arrays.

Combination of ligation and extension

In e preferred embodiment, OLA and SBE are combined, as is sometimes referred to as "Genetic Bit" analysis end described in Nikforov et al., Nucleic Acid Res. 22:4167 (1994), hereby expressly incorporated by reference. In this embodiment, the two ligation probes do not hybridize adjacently; rather, they are separated by one or more bases. The addition of dNTPs and a polymerase, in addition to the ligation probes and the ligase, results in an extended, ligated probe. As for SBE, the dNTPs may carry different labels, or separate reactions can be run, if the SBE portion of the reaction is used for genotyping. Alternatively, if the ligation portion of the reaction is used for genotyping, either no extension occurs due to mismatch of the 3' base (such that the polymerase will not extend it), or no ligation occurs due to mismatch of the 5' base. As will be appreciated by those in the art, the reaction products are essayed using microsphere arrays. Again, as outlined herein, the assays may be solution based assays, with the ligated, extended probes being added to a microsphere array, or solid-phase assays. In addition, the unextended, unligated primers may be removed prior to detection as needed, as is outlined herein. Furthermore, adapter sequences may also be used as outlined herein for OLA.

Combination of OLA and PCR

In a preferred embodiment, OLA and PCR are combined. As will be appreciated by those in the art, the sequential order of the reaction is variable. That is, in some embodiments it is desired to perform the genotyping or OLA reaction first followed by PCR amplification. In an alternative embodiment, it is

desirabla to first emplify tha target i.e. by PCR followad by the OLA assay.

In e prefarrad ambodiment, this tachniqua is dona on bead errays.

Combination of competitiva hybridlzation and ligation

In a prafarred ambodiment, a combination of competitiva hybridization and ligation is dona. As will ba appraciatad by thosa in tha art, this tachnique is OLA es described ebova, with at least two sets of probes comprising different bases in the readout position. By running the raections undar conditions that favor hybridization complaxas with parfact matchas, differant allalas may ba distinguished.

In one embodiment, LCR is used to genotype e single genomic locus by incorporating two sets of two optically labeled AS oligonucleotides end e detection oligonucleotide in the ligation raaction. The oligonucleotida ligetion stap discriminates betwaen tha AS oligonucleotides through the efficiency of ligation between an oligonucleotida with a correct match with the targat nucleic acid varsus e mismatch basa in tha targat nucleic acid et tha ligetion site. Accordingly, a detection oligonuclaotide ligates efficiently to en AS oligonucleotida if there is complete basa peiring et the ligation sita. One 3' oligonucleotida (T basa at 5' and) is opticelly labeled with FAM (green fluorascent dye) end tha othar 3' oligonucleotide (C basa at 5' and) is labelled with TMR (yellow fluorescent dye). An A base in the target nuclaic acid base pairs with the corresponding T resulting in efficiant ligation of tha FAM-labeled oligonuclaotida. A G basa in the terget nuclalc ecid rasults in ligation of the TMR-labelad oligonucleotide. TMR and FAM hava distinct emission spactrums. Accordingly, the wavelength of the oligonucleotide ligated to tha 5' detection oligonucleotide indicates the nucleotida end thus the genotypa of the target nucleic ecid.

In a preferrad embodiment, this techniqua is dona on baad arrays.

Combinetion of competitive hybridization end invasive claavage

In a preferred embodiment, a combination of compatitiva hybridization and invasiva cleavaga is dona. As will ba appreciated by those in tha ert, this technique is invasive cleavaga as described above, with at least two sets of probes (aither tha invader probes or the signalling probas) comprising different beses in the readout position. By running tha reections under conditions that favor hybridizetion complaxes with perfact matchas, diffarant allales may ba distinguished.

In a preferred embodimant, this techniqua is dona on baad errays.

In addition to tha amplification and ganotyping embodiments disclosed haraln, tha presant invantion furthar provides compositions and methods for nucleic ecid sequencing.

SEQUENCING

The present Invention is directed to the sequencing of nucleic acids, particularly DNA, by synthesizing nucleic acids using the target sequence (i.e. the nucleic acid for which the sequence Is determined) es a template. These methods can be generally described es follows. A target sequence is ettached to e solid support, either directly or indirectly, es outlined below. The target sequence comprises e first domain and an edjacent second domain comprisIng target positions for which sequence information is desired. A sequencing primer is hybridized to the first domain of the target sequence, and en extension enzyme is edded, such as a polymerase or a ligase, as outlined below. After the eddition of each base, the identity of each newly added base is determined prior to edding the next base. This can be done in e variety of ways, including controlling the reaction rate and using a fast detector, such that the newly added bases are identified in real time. Alternatively, the eddition of nucleotides is controlled by reversible chein termination, for example through the use of photocleavable blocking groups. Alternetively, the addition of nucleotides is controlled, so that the reaction is limited to one or a few bases at e time. The reaction is restarted after each cycle of addition end reading. Alternatively, the addition of nucleotides is accomplished by carrying out e ligetion reaction with oligonucleotides comprising chain terminating oligonucleotides. Preferred methods of sequencing-by-synthesis include, but are not limited to, pyrosequencing, reversible-chain termination sequencing, time-resolved sequencing, ligation sequencing, end single-molecule analysis, all of which ere described below.

The advantages of these "sequencing-by-synthesis" reactions can be eugmented through the use of arrey techniques that allow very high density errays to be made rapidly end inexpensively, thus allowing rapid and inexpensive nucleic acid sequencing. By "array techniques" is meant techniques that allow for analysis of a plurality of nucleic acids in an array format. The maximum number of nucleic acids is limited only by the number of discrete loci on a particular array platform. As is more fully outlined below, a number of different array formats can be used.

The methods of the invention find particular use in sequencing a target nucleic acid sequence, i.e. identifying the sequence of a target base or target bases in a target nucleic acid, which can ultimately be used to determine the sequence of long nucleic acids.

As is outlined herein, the target sequence comprises positions for which sequence information is desired, generally referred to herein as the "target positions". In one embodiment, a single target position is elucideted; In a preferred embodiment, a plurality of target positions ere elucidated. In general, the plurality of nucleotides in the target positions are contiguous with each other, elthough in some circumstances they may be separated by one or more nucleotides. By "plurality" as used herein is meant at least two. As used herein, the base which basepairs with the target position base in a hybrid is termed the "sequence position". That is, es more fully outlined below, the extenslon of e sequence primer results in nucleotides being added in the sequence positions, that ere perfectly complementary to the nucleotides in the target positions. As will be appreciated by one of ordinary skill in the ert, identification of a plurality of target positions in e target nucleotide sequence results in

the determination of the nucleotide sequence of the target nucleotide sequence.

As will be appreciated by one of ordinery skill in the art, this system can take on e number of different configurations, dapending on the sequancing mathod used, the method of attaching a target sequence to e surface, etc. In general, the methods of tha invention rely on the attachment of differant target sequences to a solid support (which, as outlined below, can be eccomplished in a variety of ways) to form en erray. The target sequences comprise at least two domains: a first domain, for which sequence information is not desired, and to which e sequencing primer can hybridize, and a second domain, adjacent to the first domein, comprising the target positions for sequencing. A sequencing primer is hybridized to the targat sequence, forming a hybridization complex, and than the sequencing primer is enzymatically extended by tha addition of a first nucleotide into the first sequence position of the primer. This first nucleotide is then identified, as is outlined below, and then the process is repeated, to add nucleotides to the second, third, fourth, etc. sequence positions. The exact methods depend on the sequencing technique utilized, es outlined below.

Once tha target sequance is essociated onto the array as outlined below, the target sequance can be used in a variety of sequencing by eynthesis reactions. These reactions are generally classified into severel categories, outlined below.

SEQUENCING BY SYNTHESIS

As outlinad herein, a number of sequencing by synthesis reactions are used to elucidate the identity of a plurality of basas et target positions within the target sequence. All of these reactions rely on the use of a target sequence comprising at least two domains; e first domain to which a sequencing primer will hybridiza, and en adjacent second domain, for which sequence informetion is desired. Upon formation of the assay complex, extension enzymes ere used to add dNTPs to the sequencing primer, and each addition of dNTP is "read" to determine the identity of the added dNTP. This may proceed for many cycles.

Pyrosequencing

In a preferred embodiment, pyrosequencing methods ere done to sequence tha nucleic acids. As outlined ebove, pyrosequencing is en extension method that can be used to add ona or more nucleotides to the target positions. Pyrosequencing relies on the detection of a reaction product, pyrophosphate (PPi), produced during the eddition of an NTP to a growing oligonucleotide cheln, rather than on a lebel attached to the nucleotide. One molecule of PPi is produced per dNTP added to the axtension primer. The detection of the PPi produced during the reaction is monitored using secondary enzymes; for example, preferred ambodiments utilize secondary enzymes that convert tha PPi into ATP, which elso may be detected in a variety of ways, for example through a chemiluminescent reaction using luciferese and luciferin, or by the detection of NADPH. Thus, by running sequential reactions with each of tha nucleotides, and monitoring the raaction products, the

identity of the added base is determined.

Accordingly, the present invention provides methods of pyrosequencing on arrays; the arrays may be any number of different array configurations and substrates, as outlined herein, with microsphere arrays being particularly preferred. In this embodiment, the target sequence comprises a first domain that is substantially complementary to a sequencing primer, and an adjacent second domain that comprises a plurality of target positions. By "sequencing primer" herein is meant a nucleic acid that is substantially complementary to the first target domain, with perfect complementarity being preferred. As will be appreciated by those in the art, the length of the sequencing primer will vary with the conditions used. In general, the sequencing primer ranges from about 6 to about 500 or more basepairs in length, with from about 8 to about 100 being preferred, and from about 10 to about 25 being especially preferred.

Once the sequencing primer is added and hybridized to the target sequence to form a first hybridization complex (also sometimes referred to herein as an "assay complex"), the system is ready to initiate sequencing-by-synthesis. The methods described below make reference to the use of fiber optic bundle substrates with associated microspheres, but as will be appreciated by those in the art, any number of other substrates or solid supports may be used, or arrays that do not comprise microspheres.

The reaction is initiated by introducing the substrate comprising the hybridization complex comprising the target sequence (i.e. the array) to a solution comprising a first nucleotide, generally comprising deoxynucleoside-triphosphates (dNTPs). Generally, the dNTPs comprise dATP, dTTP, dCTP and dGTP. The nucleotides may be naturally occurring, such as deoxynucleotides, or non-naturally occurring, such as chain terminating nucleotides including dideoxynucleotides, as long as the enzymes used in the sequencing/detection reactions are still capable of recognizing the analogs. In addition, as more fully outlined below, for example in other sequencing-by-synthesis reactions, the nucleotides may comprise labels. The different dNTPs are added either to separate aliquots of the hybridization complex or preferably sequentially to the hybridization complex, as is more fully outlined below. In some embodiments it is important that the hybridization complex be exposed to a single type of dNTP at a time.

In addition, as will be appreciated by those in the art, the extension reactions of the present invention allow the precise incorporation of modified bases into a growing nucleic acid strand. Thus, any number of modified nucleotides may be incorporated for any number of reasons, including probing structure-function relationships (e.g. DNA:DNA or DNA:protein interactions), cleaving the nucleic acid, crosslinking the nucleic acid, incorporate mismatches, etc.

In addition to a first nucleotide, the solution also comprises an extension enzyme, generally a DNA

polymerase. Suitable DNA polymerases include, but ere not limited to, the Klenow fragment of DNA polymerese I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemicel), T5 DNA polymerese end Phl29 DNA polymerese. If the dNTP is complementary to the base of the target sequence edjacent to the extension primer, the extension enzyme will edd it to the extension primer, releasing pyrophosphate (PPi). Thus, the extension primer is modified, i.e. extended, to form a modified primer, sometimes referred to herein as a "newly synthesized strend". The incorporetion of e dNTP into e newly synthesized nucleic ecid strand releases PPi, one molecule of PPi per dNTP incorporeted.

The release of pyrophosphate (PPi) during the DNA polymerese reaction can be quantitatively measured by many different methods end a number of enzymatic methods have been described; see Reeves et al., Anal. Biochem. 28:282 (1969); Guillory et al., Anal. Biochem. 39:170 (1971); Johnson et al., Anal. Biochem. 15:273 (1968); Cook et al., Anal. Biochem. 91:557 (1978); Drake et al., Anal. Biochem. 94:117 (1979); Ronaghi et al., Science 281:363 (1998); Bershop et al., Anal. Biochem. 197(1):266-272 (1991) WO93/23564; WO 98/28440; WO98/13523; Nyren et al., Anal. Biochem. 151:504 (1985); ell of which are incorporated by reference. The latter method allows continuous monitoring of PPi and has been termed ELIDA (Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay). In a preferred embodiment, the PPi is detected utilizing UDP-glucose pyrophosphorylase, phosphoglucomutase and glucose 6-phosphate dehydrogenese. See Justesen, et al., Anal. Biochem. 207(1):90-93 (1992); Lust et al., Clin. Chem. Acta 66(2):241 (1976); end Johnson et al., Anal. Biochem. 26:137 (1968); all of which are hereby incorporated by reference. This reaction produces NADPH which can be detected fluoremetrically.
A preferred embodiment utilizes any method which can result in the generation of an optical signal, with preferred embodiments utilizing the generation of a chemiluminescent or fluorescent signel.

Generally, these methods rely on secondary enzymes to detect the PPi; these methods generally rely on enzymes that will convert PPi into ATP, which can then be detected. A preferred method monitors the creation of PPi by the conversion of PPi to ATP by the enzyme sulfurylase, and the subsequent production of visible light by firefly luciferase (see Ronaghi et al., supra, end Barshop, supra). In this method, the four deoxynucleotides (dATP, dGTP, dCTP and dTTP; collectively dNTPs) ere added stepwise to a partial duplex comprising a sequencing primer hybridized to a single stranded DNA template and incubated with DNA polymerase, ATP sulfurylase (and its substrate, adenosine 5'-phosphosulphate (APS)) luciferase (end its substrate luciferin), and optionally e nucleotide-degrading enzyme such as apyrase. A dNTP is only incorporated into the growing DNA strend if it is complementary to the base in the template strend. The synthesis of DNA is accompanied by the release of PPi equal in molerity to the incorporated dNTP. The PPi is converted to ATP end the light generated by the luciferese is directly proportional to the amount of ATP. In some cases the unincorporated dNTPs and the produced ATP are degraded between each cycle by the nucleotide degrading enzyme.

As will be eppreciated by those in the art, if the target sequence comprises two or more of the seme nucleotide in a row, more than one dNTP will be incorporated; however, the emount of PPi generated is directly proportionel to the number of dNTPs incorporated and thus these sequences can be detected.

In eddition, in a preferred embodiment, the dATP that is added to the reaction mixture is an anelog that can be incorporated by the DNA polymerase into the growing oligonucleotide strand, but will not serve es e substrete for the second enzyme; for exemple, certain thiol-containing dATP analogs find particular use.

Accordingly, e preferred embodiment of the methods of the invention is es follows. A substrate comprising microspheres containing the target sequences end extension primers, forming hybridization complexes, is dipped or contacted with a volume (reaction chamber or well) comprising a single type of dNTP, en extension enzyme, and the reagents and enzymes necessary to detect PPi. If the dNTP is complementary to the base of the target portion of the target sequence adjacent to the extension primer, the dNTP Is added, releesing PPi and generating detectable light, which is detected as generally described in U.S.S.N.s 09/151,877 and 09/189,543, and PCT US98/09163, all of which are hereby incorporated by reference. If the dNTP is not complementary, no detectable signal results. The substrate is then contacted with a second reaction chamber comprising a different dNTP and the additional components of the assey. This process is repeated to generate a readout of the sequence of the target sequence.

In a preferred embodiment, washing steps, i.e. the use of washing chambers, may be done in between the dNTP reaction chambers, as required. These washing chambers mey optionally comprise a nucleotide-degrading enzyme, to remove any unreacted dNTP and decreasing the background signal, as is described in WO 98/28440, incorporated herein by reference. In e preferred embodiment a flow cell is used as a reaction chamber; following each reaction the unreacted dNTP is washed eway and may be replaced with en additional dNTP to be examined.

As will be appreciated by those in the art, the system can be configured in a variety of ways, including both a linear progresslon or e circular one; for example, four substrates may be used that each can dip into one of four reaction chambers arrayed in e circular pattern. Each cycle of sequencing and reading is followed by a 90 degree rotation, so that each substrate then dips into the next reaction well. This allows a continuous series of sequencing reactions on multiple substrates in perallel.

In a preferred embodiment, one or more internal control sequences are used. That is, at least one microsphere in the array comprises a known sequence that can be used to verify that the reactions ere proceeding correctly. In a preferred embodiment, at least four control sequences are used, each of which has a different nucleotide at each position: the first control sequence will have an adenosine at

position 1, the second will heve a cytosine, the third a guenosine, end the fourth a thymidine, thus ensuring that et least one control sequence is "lighting up" et eech step to serve es en internel control.

In e preferred embodiment, the reection is run for a number of cycles until the signal-to-noise ratio becomes low, generally from 20 to 70 cycles or more, with from ebout 30 to 50 being etandard. In some embodiments, this is sufficient for the purposes of the experiment; for example, for the detection of certain mutations, including single nucleotide polymorphisms (SNPs), the experiment is designed such that the initial round of sequencing gives the desired information. In other embodiments, it is desirable to sequence longer targets, for example in excess of hundreds of bases. In this epplicetion, additional rounds of sequencing can be done.

For example, efter a certain number of cycles, it is possible to stop the reaction, remove the newly synthesized strand using either a thermal step or a chemical wash, and start the reection over, using for example the sequence information thet was previously genereted to make a new extension primer that will hybridize to the first target portion of the terget sequence. That is, the sequence information generated in the first round is trensferred to an oligonucleotide synthesizer, and e second extension primer is made for e second round of sequencing. In this way, multiple overlepping rounds of sequencing are used to generate long sequences from template nucleic acid strands. Alternatively, when a single target sequence contains e number of mutational "hot spots", primers can be generated using the known sequences in between these hot spots.

Additionally, the methods of the invention find use in the decoding of random microsphere errays. That is, es described in U.S.S.N. 09/189,543, nucleic acids can be used es bead identifiers. By using sequencing-by-synthesis to read out the sequence of the nucleic ecids, the beads can be decoded in a highly parallel feshion.

In eddition, the methods find use in simultaneous enalysis of multiple target sequence positions on a single array. For example, four separete sequence analysis reections are performed. In the first reaction, positions containing a particular nucleotide ("A", for example) in the target sequence are enelyzed. In three other reactions, C, G, end T are analyzed. An edventage of anelyzing one base per reaction is that the baseline or background is flattaned for the three bases excluded from the reaction. Therefore, the signel is more eesily detected end the sensitivity of the essay is increased. Alternatively, eech of the four sequencing reactions (A, G, C and T) cen be performed simultaneously with e nested set of primers providing a significent advantage in that primer synthesis can be made more efficient.

In enother preferred embodiment each probe is represented by multiple beads in the erray (see U.S.S.N. 09/287,573, filed April 6, 1999, hereby expressly incorporated by reference) . As a result, each experiment can be repliceted many times in parallel. As outlined below, averaging the signel

from each respective probe in an experiment also allows for improved signal to noise and increases the sensitivity of detecting subtle perturbations in signal intensity patterns. The use of redundancy and comparing the patterns obtained from two different samples (e.g. a reference and an unknown), results in highly parallelled and comparative sequence analysis that can be performed on complex nucleic acid samples.

As outlined herein, the pyrosequencing systems may be configured in a variaty of ways; for example, the target sequence may be attached to the array (e.g. the beads) in a variety of ways, including the direct attachment of the target sequence to the array; the use of a capture probe with a separate extension probe; the usa of a capture extender probe, a capture probe and a seperate extension probe; the use of adapter sequences in the target sequence with capture and extension probes; and the use of a capture probe that also serves as the extansion probe.

In addition, as will be appreciated by those in the art, the target sequence may comprise any number of sets of different first and second target domains; that is, depending on the number of target positions that may be elucidated at a time, there may be several "rounds" of sequencing occuring, each time using a different target domain.

One additional benefit of pyrosequencing for genotyping purposes is that since the reaction does not rely on the incorporation of labels into a growing chain, the unreacted extension primers need not be removed.

Thus, pyrosequencing kits and reactions require, in no particularly order, arrays comprising capture probes, sequencing primers, an extension enzyme, and secondary enzymes and reactants for the detection of PPi, generally comprising enzymes to convert PPi into ATP (or other NTPs), and enzymes and reactants to detect ATP.

Attachment of enzymes to arrays

In a preferred embodiment, particularly when secondary enzymes (i.e. enzymes other than extension enzymes) are used in the reaction, the enzyme(s) may be attached, preferably through the use of flexible linkers, to the sites on the array, e.g. the beads. For example, when pyrosequencing is done, one embodiment utilizes detection based on the generation of a chemiluminescent signal in the "zone" around the bead. By attaching the secondary enzymes required to generate the signal, an increased concentration of the required enzymes is obtained in the immediate vicinity of the reaction, thus allowing for the use of less enzyme and faster reaction rates for detection. Thus, preferred embodiments utilize the attachment, preferably covalently (although as will be appreciated by those in the art, other attachment mechanisms may be used), of the non-extension secondary enzymes used to generate the signal. In some embodiments, the extension enzyme (e.g. the polymerase) may be attached as well, although this is not generally preferred.

The ettachment of enzymes to erray sites, particulerly beads, is outlined in U.S.S.N. 09/287,573, hereby incorporated by reference, end will be appreciated by those in the ert. In general, the use of flexible linkers ere preferred, es this ellows the enzymes to interect with the substrates. However, for some types of ettechment, linkers ere not needed. Attachment proceeds on the basis of the composition of the errey site (i.e. either the substrete or the beed, depending on which erray eystem is used) end the composition of the enzyme. In e preferred embodiment, depending on the composition of the erray site (e.g. the beed), it will contain chemical functional groups for subsequent ettachment of other moieties. For example, beeds comprising a variety of chemicel functionel groups such es emines ere commercially evaileble. Preferred functionel groups for ettachment are emino groups, cerboxy groups, oxo groups end thiol groups, with emino groups being particulerly preferred. Using these functional groups, the enzymes can be atteched using functionel groups on the enzymes. For example, enzymes containing amino groups cen be attached to particles comprising amino groups, for example using linkers as ere known in the art; for example, homo-or hetero-bifunctional linkers as ere well known (see 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated herein by reference).

Reversible Chain Terminetion Methods

In a preferred embodiment, the sequencing-by-synthesis method utilized is reversible chain terminetion. In this embodiment, the rete of eddition of dNTPs is controlled by using nucleotide analogs thet contain e removeble protecting group et the 3' position of the dNTP. The presence of the protecting group prevents further addition of dNTPs at the 3' end, thus ellowing time for detection of the nucleotide added (for example, utilizing a labeled dNTP). After ecquisition of the identity of the dNTP edded, the protecting group is removed and the cycle repeated. In this way, dNTPs are added one et a time to the sequencing primer to allow elucidetion of the nucleotides at the target positions. See U.S. Petent Nos. 5,902,723; 5,547,839; Metzker et al., Nucl. Acid Res. 22(20):4259 (1994); Canard et al., Gene 148(1):1-6 (1994); Dyatkina et el., Nucleic Acid Symp. Ser. 18:117-120 (1987); ell of which are hereby expressly incorporated by reference.

Accordingly, the present invention provides methods and compositions for reversible chein termination sequencing-by-synthesis. Similar to pyrosequencing, the reaction requires the hybridization of e substantielly complementary sequencing primer to a first target domeln of e target sequence to form en essey complex.

The reection is initiated by introducing the essay complex comprising the target sequence (i.e. the errey) to e solution comprising e first nucleotide enelog. By "nucleotide enelog" in this context herein is meant e deoxynucleoside-triphosphete (elso called deoxynucleotides or dNTPs, i.e. dATP, dTTP, dCTP end dGTP), thet is further derivetized to be reversibly cheln terminating. As will be eppreciated by those in the ert, any number of nucleotide enelogs may be used, as long as a polymerase enzyme will still incorporate the nucleotide et the sequence position. Preferred embodiments utilize 3'-O-

methyl-dNTPs (with photolytic removal of the protecting group), 3'-substituted-2'-dNTPs that contain anthranylic derivatives that are fluorescent (with alkali or enzymatic treatment for removal of the protecting group). The latter has the advantage that the protecting group is also the fluorescent label; upon cleavage, the label is also removed, which may serve to generally lower the background of the assay as well.

Again, the system may be configured and/or utilized in a number of ways. In a preferred embodiment, a set of nucleotide analogs such as derivatized dATP, derivatized dCTP, derivatized dGTP and derivatized dTTP is used, each with a different detectable and resolvable label, as outlined below. Thus, the identification of the base at the first sequencing position can be ascertained by the presence of the unique label.

Alternatively, a single label is used but the reactions are done sequentially. That is, the substrate comprising the array is first contacted with a reaction mixture of an extension enzyme and a single type of base with a first label, for example ddATP. The incorporation of the ddATP is monitored at each site on the array. The substrate is then contacted (with optional washing steps as needed) to a second reaction mixture comprising the extension enzyme and a second nucleotide, for example ddTTP. The reaction is then monitored; this can be repeated for each target position.

Once each reaction has been completed and the identification of the base at the sequencing position is ascertained, the terminating protecting group is removed, e.g. cleaved, leaving a free 3' end to repeat the sequence, using an extension enzyme to add a base to the 3' end of the sequencing primer when it is hybridized to the target sequence. As will be appreciated by those in the art, the cleavage conditions will vary with the protecting group chosen.

In a preferred embodiment, the nucleotide analogs comprise a detectable label as described herein, and this may be a primary label (directly detectable) or a secondary label (indirectly detectable).

In addition to a first nucleotide, the solution also comprises an extension enzyme, generally a DNA polymerase, as outlined above for pyrosequencing.

In a preferred embodiment, the protecting group also comprises a label. That is, as outlined in Canard et al., supra, the protecting group can serve as either a primary or secondary label, with the former being preferred. This is particularly preferred as the removal of the label at each round results in less background noise, less quenching and less crosstalk.

In this way, reversible chain termination sequencing is accomplished.

Time-resolved sequencing

In e preferred embodiment, time-resolved sequencing is dona. This embodiment relias on controlling the reection rata of the extension reection and/or using e fast imeging system. Besicelly, the method involves e simple axtension reection thet is eithar "slowed down", or imeged using a fast system, or both. Whet is important is that the rate of polymerizetion (extansion) is significantly slower then the rate of imege capture.

To allow for real time sequencing, parameters such es the speed of the detector (millisecond speed is preferred), and rate of polymerization will be controlled such thet tha rate of polymerizetion is significantly slower than the rate of image capture. Polymerization retes on the order of kilobases per minute (e.g. ~10 milliseconds/nucleotide), which can be edjusted, should ellow e sufficiently wida window to find conditions where the sequentiel eddition of two nucleotides can be resolved. The DNA polymerization reection, which has been studied intensively, cen eesily be reconstituted in vitro and controlled by varying e number of perameters including reection temperature and the concentration of nucleotide triphosphates.

In eddition, the polymerese cen be applied to the primer-template complex prior to initiating the reection. This serves to synchronize the reaction. Numerous polymerases ere available. Some examples include, but are not limited to polymerases with 3' to 5' exonuclease activity, other nuclease activities, polymerases with different processivity, affinities for modified end unmodified nucleotide triphosphetes, temperatura optima, stability, and the like.

Thus, in this embodiment, the reaction proceeds es outlined above. The target sequence, comprising e first domein that will hybridize to e sequencing primer end a second domein comprising e plurelity of target positions, is attached to an array es outlined below. Tha sequencing primers ere edded, along with an extension enzyme, es outlined herein, and dNTPs are added. Again, as outlined above, aither four differently labeled dNTPs may be used simultaneously or, four different sequential reactions with e single label ere done. In general, the dNTPs comprise either a primary or e secondery label, es outlined above.

In a preferred embodiment, the extension enzyme is one that is reletively "slow". This mey be eccomplishad in several weys. In one embodiment, polymerase variants ere used that have a lower polymerization rete than wild-typa enzymes. Alternatively, the reaction rate may be controlled by varying the temperatura end the concentration of dNTPs.

In e preferred embodiment, e fast (millisecond) high-sensitivity imaging system is used.

In one embodiment, DNA polymerization (extension) is monitored using light scattering, as is outlined in Johnson et al., Anal. Biochem. 136(1):192 (1984), hereby exprassly incorporated by reference.

ATTACHMENT OF TARGET SEQUENCES TO ARRAYS

As is generally described herein, there are a variety of methods that can be used to attach target aequances to the solid supports of the invention, particulerly to the microspheres that are distributed on a surface of a substrate. Most of these methods generally rely on capture probes ettached to the array. However, the attachment may be direct or indirect. Direct attachment includes those situations wherein an endogeneous portion of the target sequence hybridizes to the capture probe, or where the target sequenca has been manipuleted to contain exogeneous edapter sequences that are added to the target sequence, for example during an emplification reaction. Indirect ettachment utilizes one or more secondary probes, termed a "capture extender probe" as outlined herein.

In a preferred embodiment, direct attachment is done, as is generally depicted in Figure 1A. In this embodiment, the target sequence comprises e first target domain that hybridizes to all or part of the capture probe.

In a preferred embodiment, direct attachmant is accomplished through the use of adapters. The edapter is a chemical moiety that allows one to address the products of e reaction to e solid surface. The type of reaction includes the amplification, genotyping and sequencing reactions disclosed herein. The adapter chemical molety is independent of the reaction. Because the adapters ere independent of the reaction, sets of edapters cen be reused to create a "universal" erray that can detect a variety of products from a reaction by attaching the set of adapters that address to specific locations within the errey to different reactants.

Typically, the adepter and the capture probe on an array are binding partners, as definad herein. Although the use of other binding partners ere possible, preferred embodiments utilize nucleic ecid adapters that ara non-complementary to any reactants or target sequances, but are substantielly complementary to all or part of the capture probe on the array.

Thus, an "adapter sequence" is e nucleic acid that is generally not native to the target sequence, i.e. is exogeneous, but is added or attached to the target sequence. It should be noted that in this context, the "target sequence" can include the primary sample target sequence, or can be a derivative target such as a reactant or product of the reections outlined herein; thus for example, the target sequence can be a PCR product, a first ligation probe or a ligated probe in an OLA reaction, etc.

As will be appreciated by those in the art, the attachment, or joining, of the adapter sequence to the target sequence can be done in a variety of ways. In a preferred embodiment, the adapter sequences ere added to the primers of the reaction (extension primers, amplification primers, readout probas, sequencing primers, Rolling Circle primers, etc.) during the chemical synthesis of the primers. The adapter then gets edded to the reaction product during the reaction; for exemple, the primer gets extendad using a polymerase to form the new target sequence that now contains an adapter

sequence. Alternatively, the adapter sequences can be added enzymatically. Furthermore, the edapter can be attached to the target after synthesis; this post-synthesis attachment could be either covalent or non-covalent.

In this embodiment, one or more of the amplification primers comprises e first portion comprising the adapter sequence and e second portion comprising the primer sequence. Extending the amplification primer as is well known in the ert results in target sequences that comprise the adapter sequences. The edapter sequences are designed to be substantielly complementary to capture probes.

In addition, as will be eppreciated by those in the art, the edepter can be attached either on the 3' or 5' ends, or in an internal position. For example, the adapter may be the detection saquence of an invasive cleavage probe. In the case of Rolling Circle probes, the adapter can be contained within the section between the probe ends. Adapters can elso be attached to aptamers. Aptamers are nucleic acids that can be made to bind to virtually any target analyte; see Bock et al., Nature 355:564 (1992); Femulok et al., Current Op. Chem. Biol. 2:230 (1998); and U.S. Patents 5,270,163, 5,475,096, 5,567,588, 5,595,877, 5,637,459, 5,683,687,5,705,337, end related patents, hereby incorporated by reference. In addition, es outlined below, the adapter can be attached to non-nucleic acid target analytes as well.

In one embodiment, a set of probes is hybridized to a target sequence; each probe is complementary to a different reglon of e single target but each contains the same adapter. Using a poly-T bead, the mRNA target is pulled out of the sample with the probes attached. Dehybridizing the probes ettached to the target sequence end rahybridizing them to an array containing tha capture probes complementary to the adepter sequences results in binding to the errey. All adapters that have boudn to the same target mRNA will bind to the same location on the array.

In a preferred embodiment, indirect attachment of the target sequence to the array is done, through the use of capture extender probes. "Capture extender" probes are generally depicted in Figure 1C, and other figures, end have a first portion that will hybridize to all or part of the capture probe, and a second portion that will hybridize to e first portion of the target sequence. Two captura extender probes may also be used. This hes generally been done to stabilize assay complexes for example when the target eaquence is large, or when lerge amplifier probes (particularly branched or dendrimer amplifier probes) are used.

When only cepture probes ere utilized, it is necessary to have unique capture probes for each target sequence; that is, the surface must be customized to contain unique capture probes; e.g. each bead comprises a different capture probe. In general, only e single type of capture probe should be bound to a bead; however, different beads should contain different capture probes so that different target sequences bind to different beads.

Alternatively, the use of adepter sequences and capture extender probes ellow the creation of more "universal" surfaces. In a preferred embodiment, en array of different and usually ertificial capture probes are made; that is, the capture probes do not have complementarity to known target sequences. Tha adaptar sequences can then ba added to any target sequances, or solubla captura axtandar probes ere made; this ellows the manufacture of only one kind of erray, with the user able to customize the erray through the use of adapter sequences or capture extender probes. This then allows the generation of customized soluble probes, which es will be appreciated by those in the ert is ganarally simpler and lass costly.

As will be appreciated by those in the art, the length of the edapter sequences will vary, depending on the desired "strength" of binding end the number of different adepters desired. In a preferred embodiment, edapter sequences renge from about 6 to about 500 basepairs in length, with from about 8 to about 100 being preferred, and from about 10 to about 25 being perticularly preferred.

In one embodiment, microsphere arrays containing a single type of capture probe are mede; in this embodiment, the capture extender probes ere added to the beeds prior to loading on the array. The capture extendar probas may be additionally fixad or crosslinkad, as necessary.

In a preferred embodiment, as outlined in Figure 1B, the capture probe comprises the sequencing primer; that is, after hybridization to the target sequence, it is the capture probe itself that is extended during the synthesis reaction.

In one embodiment, capture probes are not used, end the target sequences are attached directly to the sites on the array. For example, libraries of clonal nucleic acids, including DNA and RNA, are usad. In this ambodimant, individual nucleic acids are prapared, generally using conventionel methods (including, but not limited to, propagation in plasmid or phage vectors, amplification techniques including PCR, etc.). The nucleic ecids are preferably errayad in some format, such as e microtiter plate format, and either spotted or beeds are added for attachment of the libraries.

Attachment of the clonal libraries (or eny of the nucleic ecids outlined herein) may be done in a variety of ways, es will be appreciated by those in the art, including, but not limited to, chemical or affinity capture (for example, including the incorporation of derivatized nucleotides such es AminoLink or biotinylated nuclaotides that can then be usad to attach the nucleic acid to a surfaca, as well as effinity capture by hybridization), cross-linking, and electrostatic attachment, etc.

In a preferred embodiment, affinity capture is used to attach the clonal nucleic acids to the surface. For exampla, cloned nuclaic acids can be derivatized, for example with ona mamber of a binding pair, end the beads derivetized with the other member of a binding pair. Suitable binding pairs are es described hereln for secondary labels and IBL/DBL pairs. For example, the cloned nucleic ecids may

be biotinyleted (for example using enzymatic incorporeta of biotinylated nucleotides, for by photoactiveted cross-linking of biotin). Biotinylated nucleic ecids can then be captured on streptavidin-coeted beeds, as is known in the ert. Similerly, other hepten-receptor combinations can be used, such as digoxigenin and anti-digoxiganin antibodies. Altarnatively, chamicel groups can be added in the form of derivatized nucleotides, that can them be used to add the nucleic acid to the surface.

Preferred ettechments are covalent, elthough even relatively weak interactions (i.e. non-covalent) cen be sufficient to ettach e nucleic ecid to e surfece, if there ere multiple sites of attachment per eech nucleic ecid. Thus, for example, electrostatic interactions cen be used for attachment, for example by having beeds carrying the opposite cherge to the bioactive egent.

Similarly, affinity capture utilizing hybridization can be used to ettach cloned nucleic acids to beeds. For example, as is known in the art, polyA+RNA is routinely captured by hybridizetion to oligo-dT beeds; this may include oligo-dT capture followed by a cross-linking step, such as psorelen crosslinking). If the nucleic ecids of interest do not contain a polyA tract, one can be ettached by polymerizetion with terminal transferase, or vie ligation of en oligoA linker, as is known in the art.

Alternetively, chemical crosslinking may be done, for exemple by photoactivated crosslinking of thymidine to reective groups, es is known in the ert.

In general, special mathods are required to decode clonal errays, as is more fully outlined below.

ASSAY AND ARRAYS

All of the above compositions and methods ere directed to the detection and/or quantificetion of the products of nucleic acid reactions. The detection systems of the present invention are based on the incorporation (or in some cases, of the deletion) of a detecteble label into an assay complex on en erray.

Accordingly, the present invention provides methods and compositions useful in the detection of nucleic acids. As will be eppreciated by those in the art, the compositions of the invention can take on a wide variety of configuretions, es is generelly outlined in the Figures. As is more fully outlined below, preferred systems of the invention work es follows. A target nucleic ecid sequence is ettached (via hybridization) to en erray site. This attachment can be either directly to a capture probe on the surfaca, through the usa of adapters, or indiractly, using captura extendar probas as outlinad harain. In some embodiments, the target sequence itself comprises the labels. Alternatively, a iebel probe is then edded, forming en assey complex. The ettachment of the label probe may be direct (i.e. hybridizetion to e portion of the target sequence), or indirect (i.e. hybridizetion to en emplifier probe thet hybridizes to the target sequence), with all the required nucleic acids forming en assay complex.

Accordingly, the present invantion provides array compositions comprising at least a first substrate with a surface comprising individual sites. By "array" or "biochip" herein is meant a piurality of nucleic acids in an array format; the size of the array will depend on the composition end end usa of the array. Nucleic ecids arrays are known in the art, end can be classified in e number of ways; both ordered arrays (e.g. the ability to resolve chemistries et discreta sites), and random arrays are incluided. Ordered arrays include, but are not iimited to, those made using photolithography techniques (Affymetrix GeneChip™), spotting techniques (Synteni and others), printing techniques (Hewlett Packard and Rosetta), three dimensionel "gel pad" arrays, etc. A preferred embodiment utilizes microspheres on a veriety of substrates including fiber optic bundies, es are outlined in PCTs US98/21193, PCT US99/14387 and PCT US98/05025; WO98/50782; and U.S.S.N.s 09/287,573, 09/151,877, 09/256,943, 09/316,154, 60/119,323, 09/315,584; eii of which ere expressly incorporated by reference. While much of the discussion below is directed to the use of microsphere arrays on fiber optic bundles, any array format of nucleic acids on solid supports mey be utilized.

Arrays containing from about 2 different bioactive egents (e.g. different beads, when beads are used) to many millions can be made, with very large arrays being possible. Generally, the array will comprise from two to as many as a billion or more, depending on the size of the beads and the substrate, as well as the end use of tha array, thus very high density, high density, moderate density, low density end very low density arrays may be made. Preferred ranges for very high density arrays are from about 10,000,000 to about 2,000,000,000, with from about 100,000,000 to about 1,000,000,000 being preferred (aii numbers being in square cm). High density arrays range about 100,000 to about 10,000,000, with from about 1,000,000 to about 5,000,000 being particularly prefarred. Moderate density errays range from about 10,000 to about 100,000 being particularly preferrad, and from about 20,000 to about 50,000 being especially preferred. Low density errays are generally iess than 10,000, with from about 1,000 to about 5,000 being preferred. Very iow density errays are iess than 1,000, with from about 10 to ebout 1000 being preferred, and from about 100 to about 500 being particularly preferred. In some embodiments, the compositions of the invention may not be in array format; that is, for some embodiments, compositions comprising a singie bioactive egent may be made as well. In eddition, in some arrays, multiple substrates mey be used, either of different or identical compositions. Thus for example, large arrays may comprise a plurality of smaiier substrates.

In addition, one advantage of the present compositions is that particularly through the use of fiber optic technology, extremely high dansity arreys can be made. Thus for example, because beads of 200 µm or iess (with beads of 200 nm possible) can be used, and very small fibers are known, it is possible to have as many as 40,000 or more (in some instances, 1 million) different elements (e.g. fibers and beads) in a 1 mm^2 fiber optic bundie, with densities of greater than 25,000,000 individual beads and fibers (egeln, in some instancas as many as 50-100 million) per 0.5 cm^2 obtaineble (4 million per squere cm for 5 µ center-to-center and 100 million per square cm for 1 µ center-to-center).

By "substrate" or "solid support" or other grammatical equivalents herein is meent any material that can be modified to contain discrete individual sites appropriate for the attachment or association of beads and is amenebie to at least one datection method. As will be appreciated by those in the art, the number of possible substrates is very large. Possible substrates include, but era not limited to, glass end modifiad or functionalized gless, plastics (including ecrylics, polystyrene and copolymers of styrene and other meterials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon, etc.), polysaccherides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorgenic glasses, plastics, opticel fiber bundles, end e variety of other polymers. In general, the substrates allow optical detection and do not themselves appreciebly fluoresce.

Ganerally the substrate is flat (planar), although as will be appreciated by thosa in the ert, othar configurations of substrates may be used es well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample eccess to the beads and using a confocel microscope for detection. Similarly, the beads mey be placed on the inside surfaca of a tube, for flow-through sample analysis to minimize sample volume. Preferred substretes include optical fiber bundles as discussed below, and flat planar substrates such es glass, polystyrene and other plastics and acrylics.

In e preferred embodiment, the substrate is an optical fiber bundle or arrey, as is generally described in U.S.S.N.s 08/944,850 and 08/519,062, PCT US98/05025, and PCT US98/09163, ell of which are expressly incorporeted herein by reference. Preferred embodiments utilize preformed unitary fiber optic arrays. By "preformed unitary fiber optic arrey" herein is meant an arrey of discrete individual fiber optic strands that are co-axially disposed end joined elong their lengths. The fiber strends are generally individually cled. However, ona thing that distinguished e preformed unitary array from other fiber optic formats is that the fibers are not individually physically manipulatable; that is, one strand generally cannot be physically separeted at eny point along its length from another fibar strand.

Generally, the array of array compositions of the invention can be configured in several ways; see for example U.S.S.N. 09/473,904, hereby expressly incorporated by reference. In a preferred embodiment, as is more fully outlined below, a "one component" systam is used. Thet is, a first substrate comprising a plurelity of essay locations (sometimes elso referred to herein as "assay wells"), such es a microtiter plete, is configurad such that each essay location contains an individuel array. Thet is, the essay location and the array location ere the seme. For example, the plastic meterial of the microtiter plate cen be formed to contain e plurality of "bead wells" in the bottom of each of the assay wells. Beads containing the capture probes of the invention can then be loaded into the beed wells in each assay location as is more fully described below.

Alternatively, a "two component" system can be used. In this embodiment, the individual arrays are

formed on a second substrate, which then can be fitted or "dipped" into the first microtiter plate substrate. A preferred embodiment utilizes fiber optic bundles as the individual arrays, generally with "bead wells" etched into one surface of each individual fiber, such that the beads containing the capture probes are loaded onto the end of the fiber optic bundle. The composite array thus comprises a number of individual arrays that are configured to fit within the wells of a microtiter plate.

By "composite array" or "combination array" or grammatical equivalents herein is meant a plurality of individual arrays, as outlined above. Generally the number of individual arrays is set by the size of the microtiter plate used; thus, 96 well, 384 well and 1536 well microtiter plates utilize composite arrays comprising 96, 384 and 1536 individual arrays, although as will be appreciated by those in the art, not each microtiter well need contain an individual array. It should be noted that the composite arrays can comprise individual arrays that are identical, similar or different. That is, in some embodiments, it may be desirable to do the same 2,000 assays on 96 different samples; alternatively, doing 192,000 experiments on the same sample (i.e. the same sample in each of the 96 wells) may be desirable. Alternately, each row or column of the composite array could be the same, for redundancy/quality control. As will be appreciated by those in the art, there are a variety of ways to configure the system. In addition, the random nature of the arrays may mean that the same population of beads may be added to two different surfaces, resulting in substantially similar but perhaps not identical arrays.

At least one surface of the substrate is modified to contain discrete, individual sites for later association of microspheres. These sites may comprise physically altered sites, i.e. physical configurations such as wells or small depressions in the substrate that can retain the beads, such that a microsphere can rest in the well, or the use of other forces (magnetic or compressive), or chemically altered or active sites, such as chemically functionalized sites, electrostatically altered sites, hydrophobically/ hydrophilically functionalized sites, spots of adhesive, etc.

The sites may be a pattern, i.e. a regular design or configuration, or randomly distributed. A preferred embodiment utilizes a regular pattern of sites such that the sites may be addressed in the X-Y coordinate plane. "Pattern" in this sense includes a repeating unit cell, preferably one that allows a high density of beads on the substrate. However, it should be noted that these sites may not be discrete sites. That is, it is possible to use a uniform surface of adhesive or chemical functionalities, for example, that allows the attachment of beads at any position. That is, the surface of the substrate is modified to allow attachment of the microspheres at individual sites, whether or not those sites are contiguous or non-contiguous with other sites. Thus, the surface of the substrate may be modified such that discrete sites are formed that can only have a single associated bead, or alternatively, the surface of the substrate is modified and beads may go down anywhere, but they end up at discrete sites.

In a preferred embodiment, the surface of the substrate is modified to contain wells, i.e. depressions in the surface of the substrate. This may be done as is generally known in the art using a variety of

techniquas, including, but not limited to, photolithography, stamping techniques, molding tachniquas end microetching techniquas. As will be appreciated by those in tha art, the technique used will depend on the composition end shepe of the substrate.

In a preferred embodimant, physical alterations are meda in a surface of the substrate to produce the sites. In a prefarred embodimant, tha substrate is a fiber optic bundle end tha surface of the substrate is e terminal end of the fibar bundla, as is ganerally described in 08/818,199 end 09/151,877, both of which ara hereby expressly incorporated by rafaranca. In this embodiment, wells ere made in a terminal or distal end of a fiber optic bundle comprising individual fibers. In this embodimant, the cores of the individuel fibers ere etchad, with respect to the cladding, such that small wells or depressions are formed et one end of the fibars. The required depth of the wells will depend on the size of tha beeds to be edded to the wells.

Ganerally in this ambodiment, the microsphares ara non-covalantly associated in the wells, eithough the wells may additionally ba chemically functionalized as is generally described below, cross-linking agents may ba used, or a physical barrier may be used, i.e. a film or membrane over the beads.

In a preferred embodiment, the surface of the substrate is modified to contain chemically modified sites, that can ba used to ettach, either covalently or non-covalently, the microspheres of the invention to tha discrata sites or locations on the substrata. "Chemically modified sites" in this context includes, but is not limited to, the eddition of e pattam of chemical functionel groups including amino groups, carboxy groupe, oxo groups end thiol groups, thet cen be used to covalently attach microspheres, which genarally elso conteln corresponding reective functional groups; the addition of a pattern of adhesive thet cen ba used to bind the microspheres (eithar by prior chemical functionalization for the addition of the edhesive or direct addition of the adhesive); the eddition of e pettern of charged groups (similar to tha chemicel functionalities) for the electrostatic attachment of the microsphares, i.e. when the microspheres comprisa chargad groups opposita to tha sites; the addition of a pattern of chamical functional groups that rendars the sites differantially hydrophobic or hydrophilic, such that the addition of similarly hydrophobic or hydrophilic microspheres under suitable experimental conditions will result in association of the microsphares to the sites on the basis of hydroaffinity. For exampla, tha use of hydrophobic sites with hydrophobic baads, in en equeous system, drives the associetion of the beads preferantially onto tha sites. As outlined abova, "pattern" in this sense includes the use of a uniform treatment of the surface to allow attachment of the beads at discreta sites, as well as treetment of the surface resulting in discrete sites. As will be appraciated by those in tha art, this may be eccomplished in a variety of ways.

In some embodimants, the beads ere not associated with a substrate. That is, the baads are in solution or ere not distributed on a patterned substrete.

In a preferred embodiment, the compositions of the invention further comprise a population of microspheres. By "population" herein is meant a plurality of beads as outlined above for arrays. Within the population are separate subpopulations, which can be a single microsphere or multiple identical microspheres. That is, in some embodiments, as is more fully outlined below, the array may contain only a single bead for each capture probe; preferred embodiments utilize a plurality of beads of each type.

By "microspheres" or "beads" or "particles" or grammatical equivalents herein is meant small discrete particles. The composition of the beads will vary, depending on the class of capture probe and the method of synthesis. Suitable bead compositions include those used in peptide, nucleic acid and organic moiety synthesis, including, but not limited to, plastics, ceramics, glass, polystyrene, methylstyrene, acrylic polymers, paramagnetic materials, thoria sol, carbon graphite, titanium dioxide, latex or cross-linked dextrans such as Sepharose, cellulose, nylon, cross-linked micelles and Teflon may all be used. *"Microsphere Detection Guide"* from Bangs Laboratories, Fishers IN is a helpful guide.

The beads need not be spherical; irregular particles may be used. In addition, the beads may be porous, thus increasing the surface area of the bead available for either capture probe attachment or tag attachment. The bead sizes range from nanometers, i.e. 100 nm, to millimeters, i.e. 1 mm, with beads from about 0.2 micron to about 200 microns being preferred, and from about 0.5 to about 5 micron being particularly preferred, although in some embodiments smaller beads may be used.

It should be noted that a key component of the invention is the use of a substrate/bead pairing that allows the association or attachment of the beads at discrete sites on the surface of the substrate, such that the beads do not move during the course of the assay.

Each microsphere comprises a capture probe, although as will be appreciated by those in the art, there may be some microspheres which do not contain a capture probe, depending on the synthetic methods.

Attachment of the nucleic acids may be done in a variety of ways, as will be appreciated by those in the art, including, but not limited to, chemical or affinity capture (for example, including the incorporation of derivatized nucleotides such as AminoLink or biotinylated nucleotides that can then be used to attach the nucleic acid to a surface, as well as affinity capture by hybridization), cross-linking, and electrostatic attachment, etc. In a preferred embodiment, affinity capture is used to attach the nucleic acids to the beads. For example, nucleic acids can be derivatized, for example with one member of a binding pair, and the beads derivatized with the other member of a binding pair. Suitable binding pairs are as described herein for IBL/DBL pairs. For example, the nucleic acids may be biotinylated (for example using enzymatic incorporate of biotinylated nucleotides, for by

photoactivated cross-linking of biotin). Biotinyleted nucleic acids can then be captured on streptavidin-coated beads, es is known in the art. Similarly, other hapten-receptor combinations can be used, such as digoxigenin and anti-digoxigenin entibodies. Altemetively, chemical groups can be added in the form of derivatized nucleotides, that can them be used to add the nucleic acid to the surface.

Preferred attachments ere covalent, elthough even relatively weak interactions (i.e. non-covalent) can be sufficient to ettach e nucleic ecid to a surface, if there ere multiple sitas of ettachment per eech nucleic acid. Thus, for example, electrostatic interactions can be used for attachment, for example by having beeds carrying the opposita cherga to the bioactive agent.

Similarly, effinity capture utilizing hybridization can be used to ettach nucleic acids to beeds. For example, es is known in the art, polyA+RNA is routinely captured by hybridization to oligo-dT beads; this may include oligo-dT capture followed by a cross-linking step, such as psoralen crosslinking). If the nucleic ecids of interest do not contain a polyA tract, one can be attached by polymerization with terminal transferase, or via ligation of an oligoA linker, as is known in the art.

Alternatively, chemical crosslinking mey ba done, for example by photoactivated crosslinking of thymidine to reactive groups, as is known in the art.

In generel, probes of the present invention are designed to be complementary to e targat sequance (either the target sequence of the semple or to other probe sequences, as is described herein), such that hybridization of the target and tha probes of the present invention occurs. This complementerily need not be perfect; there mey be eny number of base pair mismatches that will interfere with hybridization between the terget sequence end the single stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantielly complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under the selected reaction conditions.

In a preferred ambodiment, each beed comprises a single type of capture probe, although a plurality of individual capture probes are preferably attached to eech bead. Similarly, preferred embodiments utilize more than one microsphere containing a unique capture probe; that is, there is redundancy built into the system by the use of subpopulations of microspheres, each microsphera in the subpopulation containing the same capture probe.

As will be eppreciatad by those in the art, the capture probes may either be synthesized diractly on the beeds, or they may be mada and then ettached after synthesis. In e preferred embodiment, linkers are used to attach the capture probes to the beeds, to allow both good attachmant, sufficient flexibility to allow good interaction with the targat molecule, and to avoid undesirable binding reactions.

In e preferred embodiment, the capture probes are synthesized directly on the beads. As is known in tha art, many classes of chemical compounds ara currantly synthasized on solid supports, such as peptides, organic moieties, and nucleic ecids. It is a relatively straightforward matter to adjust the currant synthetic techniquas to usa beads.

In a preferred embodiment, the capture probes are synthesized first, and then covalently attached to the beads. As will be appreciated by those in the art, this will be done depending on the composition of the capture probas and the beads. The functionalization of solid support surfaces such as certain polymers with chemically reactive groups such as thiols, amines, carboxyls, etc. is generally known in the ert. Accordingly, "blank" microspheres may be used thet have surface chemistries that facilitate the attachment of the desired functionality by the user. Some examples of these surface chemistries for blenk microspheras include, but are not limited to, amino groups including aliphatic and aromatic amines, cerboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates.

When rendom errays are used, an encoding/decoding system must be usad. For exampe, when microsphere arrays are used, the beads are generally put onto the substrate randomly; es such there are several ways to correlate the functionality on the bead with its location, including tha incorporation of unique optical signatures, generally fluorescent dyes, that could be used to identify the chemical functionality on any particular bead. This allows the synthesis of the cendidete agents (i.e. compounds such es nucleic acids and antibodias) to be divorcad from thair placament on an array, i.a. the candidate agents may be synthesized on the beads, end then the beads are randomly distributed on a pattarnad surface. Since the beads are first coded with an optical signature, this maans thet the array cen later be "decoded", i.e. after the array is made, a correlation of the locetion of an individual site on the array with tha bead or candidete agant at that particular sita can ba mada. This means that tha beeds may be rendomly distributed on the array, a fast and inexpensive process es compared to aithar the in situ synthasis or spotting techniques of the prior art.

However, the drewback to these methods is that for a lerge array, the system requires e large number of different optical signatures, which may be difficult or time-consuming to utilize. Accordingly, the present invention provides several improvements over these methods, generally directed to methods of coding and decoding the arreys. That is, as will be appreciated by those in the art, the placement of the capture probes is generally random, and thus e coding/decoding system is required to identify the probe et each locetion in the array. This may ba done in a variety of ways, as is more fully outlined below, and generally includes: e) the use a decoding binding ligand (DBL), genarelly directly labeled, that binds to either the capture probe or to identifier binding ligands (IBLs) attached to the beads; b) positional decoding, for exemple by either targeting the placement of beads (for example by using photoactivatible or photocleavable moieties to ellow the selective eddition of beads to particular locations), or by using either sub-bundles or selective loading of the sites, as are more fully outlined

below; c) selective decoding, wherein only those beads that bind to a target are decoded; or d) combinations of eny of these. In some cases, es is more fully outlined below, this decoding may occur for all the beads, or only for those that bind e particular target sequence. Similarly, this may occur either prior to or efter eddition of e target sequence. In addition, as outlined herein, the target sequences detected may be either a primary target sequence (e.g. a patient sample), or e reaction product from one of the methods described herein (e.g. an extended SBE probe, a ligated probe, a cleaved signal probe, atc.).

Once the identity (i.e. the ectual agent) and location of each microsphere in the array hes been fixed, tha array is exposed to samples containing tha targat saquances, although as outlinad below, this can ba done prior to or during the enalysis as well. The targat sequences can hybridize (either directly or indirectly) to the capture probes as is more fully outlined below, end results in a change in the opticel signal of a particular bead.

In the present invention, "decoding" does not rely on the use of optical signatures, but rather on the use of decoding binding ligands that are added during a decoding step. The decoding binding ligends will bind either to a distinct identifier binding ligand partner that is placed on the beads, or to the capture probe itself. The decoding binding ligands are either directly or indirectly labeled, and thus dacoding occurs by detecting the presence of the label. By using pools of decoding binding ligands in e sequential fashion, it is possible to greatly minimize the number of required decoding steps.

In aome embodiments, the microspheres mey edditionally comprise identifier binding ligands for use in cartain decoding systems. By "identifier binding ligends" or "IBLs" herein is meant e compound that will specifically bind a corresponding dacoder binding ligand (DBL) to facilitate the alucidation of the identity of the capture probe attached to the bead. That is, the IBL end the corresponding DBL form a binding partner peir. By "specifically bind" herein is meant that the IBL binds its DBL with specificity sufficient to differentiate batween tha corresponding DBL and other DBLs (that is, DBLs for other IBLs), or other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the decoding step, including wash steps to remove non-specific binding. In some embodiments, for example when the IBLs and corresponding DBLs ere proteins or nucleic acids, the dissociation constants of the IBL to its DBL will be less than about 10^{-4}-10^{-6} M^{-1}, with less than ebout 10^{-5} to 10^{-9} M^{-1} being preferred and less than about 10^{-7} -10^{-9} M^{-1} being particularly preferred.

IBL-DBL binding pairs ere known or can be reedily found using known techniques. For example, when the IBL is e protein, the DBLs include proteins (particularly including antibodies or fragments thereof (FAbs, etc.)) or small molecules, or vice versa (the IBL is an antibody and the DBL is a protein). Metal ion- metal ion ligands or chelators pairs are also usaful. Antigen-antibody pairs, enzymes and substrates or inhibitors, other protein-protein interacting pairs, receptor-ligands, complementary

nucleic acids, and carbohydrates end their binding partners ere also suitable binding pairs. Nucleic acid - nucleic ecid binding proteins pairs ere elso usaful. Similerly, as is generelly described in U.S. Petents 5,270,163, 5,475,096, 5,567,588, 5,595,877, 5,637,459, 5,683,867,5,705,337, end releted patents, hareby incorporated by raference, nucleic ecid "aptamers" cen be developed for binding to virtuelly eny target; such en aptamer-target pair cen be used es the IBL-DBL pair. Simllarly, thara is a wide body of literature relating to the development of binding pairs based on combinatorial chemistry methods.

In a preferred embodimant, the IBL is a molecule whose color or luminescence properties chenge in the presence of a selectively-binding DBL. For axemple, the IBL may be e fluorescent pH indicator whosa emission intensity changes with pH. Similarly, the IBL mey be a fluorescent ion indicator, whosa emission properties chenge with ion concentration.

Alternatively, the IBL is a molecule whose color or luminescence properties change in the presence of various solvents. For example, the IBL may be a fluorescent molecule such as an ethidium selt whose fluorescence intensity increases in hydrophobic environmants. Similarly, the IBL may be a derivative of fluorescein whose color changes between aqueous and nonpolar solvents.

In one embodiment, the DBL may be etteched to e beed, i.e. e "decoder bead", that may carry a label such as e fluorophore.

In e preferred embodiment, the IBL-DBL peir comprise substantially complementary single-stranded nucleic ecids. In this embodiment, the binding ligands can be referred to es "identifier probes" and "decoder probes". Generelly, the identifier and decoder probas renge from ebout 4 besepairs in length to ebout 1000, with from ebout 6 to ebout 100 baing preferred, end from ebout 8 to about 40 beIng particulerly preferred. Whet is important is thet the probes ere long enough to be specific, i.e. to distinguish between different IBL-DBL pairs, yet short enough to ellow both e) dissociation, if necessary, under suitable experimental conditions, and b) efficient hybridization.

In e preferred embodiment, as is more fully outlined below, the IBLs do not bind to DBLs. Rather, the IBLs are used as identifier moleties ("IMs") that ere identified directly, for exampla through the use of mess spectroscopy.

Alternatively, in a praferred embodiment, the IBL end the capture probe ere the same molety; thus, for example, as outlined herein, particularly when no optical signatures ere used, tha capture probe can serve es both the identifiar end the egent. For example, in the casa of nucleic acids, the bead-bound probe (which sarves es the capture probe) cen elso bind decoder probes, to identify the sequence of the probe on tha bead. Thus, in this embodiment, the DBLs bind to the capture probes.

In e preferred ambodiment, tha microspheres mey contain an optical signature. Thet is, as outlinad in U.S.S.N.s 08/818,199 and 09/151,877, previous work had eech subpopulation of microspheres comprising a uniqua optical signature or optical tag thet is used to idantify the unique ceptura probe of thet subpopuletion of microspharas; that is, decoding utilizes optical properties of the beeds such that e bead comprising the uniqua optical signeture may be distinguished from beads at other locations with different opticel signatures. Thue the previoue work assigned eech captura probe e unique optical signatura such that eny microspheres comprising thet ceptura probe ere idantifiable on tha besis of the signature. These optical signatures comprised dyes, usuelly chromophores or fluorophores, that ware entrapped or ettached to the beeds themselves. Diversity of optical signetures utilized diffarant fluorochromes, differant ratios of mixtures of fluorochromes, end different concentrations (intansities) of fluorochromas.

In a preferred embodiment, the present invention does not rely solely on tha usa of optical properties to decode tha errays. However, es will be apprecieted by those in the ert, it is possibla in some embodiments to utilize optical signatures as an edditional coding method, in conjunction with the present system. Thus, for example, ee is mora fully outlined below, the siza of the array mey be effectively increased whila using e singla set of decoding moieties in severel ways, ona of which is the use of optical signatures one some beads. Thus, for example, using one "sat" of decoding molecules, the use of two populations of beads, one with en optical signatura and one without, allows tha effective doubling of the array size. Tha usa of multiple opticel signatures similarly increases the possible size of the array.

In a preferred embodiment, each subpopuletion of beads comprises e plurelity of different IBLs. By using e plurality of diffarent IBLs to encode eech capture probe, the number of possible unique codes is substentielly increesed. That is, by using ona unique IBL per cepture probe, the siza of tha erray will be the number of unique IBLs (assuming no "reuse" occurs, as outlined below). However, by using a plurality of different IBLs per beed, n, the size of the array can be increesed to 2^n, when the presenca or ebsenca of eech IBL is used as tha indicator. For example, tha essignmant of 10 IBLs per beed generetas e 10 bit binery code, where each bit cen be designeted es "1" (IBL is present) or "0" (IBL is absent). A 10 bit binery code hes 2^{10} possibla variants However, as is more fully discussed below, the size of the array mey be further incraased if another parameter is included euch as concentration or intensity; thus for axample, if two diffarent concentrations of the IBL are used, then the array size increases es 3^n. Thus, in this ambodiment, aach individual cepture probe in the erray is assigned a combinetion of IBLs, which can ba edded to the beads prior to tha eddition of the cepture probe, after, or during tha synthasis of the captura proba, i.e. simultaneous addition of IBLs and capture probe components.

Alternatively, tha combination of different IBLs cen be used to elucidate the sequence of the nucleic acid. Thus, for example, using two different IBLs (IBL1 end IBL2), the first position of a nucleic acid

can ba alucideted: for exemple, edenosine can be reprasanted by tha prasance of both IBL1 and IBL2; thymidine cen be representad by the presence of IBL1 but not IBL2, cytosine cen be represented by tha presence of IBL2 but not IBL1, and guanosine can be represented by the absence of both. The second position of tha nucleic acid can be done in e similer menner using IBL3 end IBL4; thus, the presence of IBL1, IBL2, IBL3 and IBL4 gives a sequanca of AA; IBL1, iBL2, and IBL3 shows tha sequence AT; IBL1, IBL3 end iBL4 gives the sequence TA, etc. The third position utilizes IBL5 end IBL6, etc. In this way, the use of 20 different identifiers cen yield e unique code for every possible 10-mer.

In this way, e sort of "bar code" for eech sequence can be constructed; the presence or ebsence of each distinct IBL will ellow the identificetion of each cepture probe.

In addition, tha use of diffarent concentrations or dansitias of IBLs allows a "reuse" of sorts. If, for example, the beed comprising e first egent has e 1X concentration of IBL, and e second bead comprising e second agent has e 10X concentration of IBL, using saturating concentrations of the correeponding lebelled DBL allows the user to distinguish between the two beads.

Once the microspheres comprising the capture probes are generated, they ere added to the substrate to form an array. It should be noted that while most of the methods described herein edd the beeds to the substrate prior to the assay, the order of making, using end decoding the errey cen very. For axample, the erray cen be made, decoded, and then the essay done. Alternetively, the array cen be made, used in en essay, and then decoded; this may find particular use when only e few beads need be decoded. Altematively, the beads can be added to the essay mixture, i.e. the semple containing the target sequences, prior to the eddition of the beeds to the substrate; after eddition and assey, the array may be dacoded. This is particularly preferred whan the sampla comprising the beads is egiteted or mixed; this can increase the amount of target sequence bound to the beads per unit time, end thus (in the case of nucleic acid essays) increase the hybridizetion kinetics. This may find particular use in cases where the concentration of target sequence in the sample is low; generally, for low concentretions, long binding times must be used.

In general, the methods of meking the arrays end of decoding the errays is done to maximize the number of different candidete egents that can be uniquely encoded. The compositions of the invention may be made in a variety of ways. In general, the arrays ara made by adding a solution or slurry comprising the beeds to e surfece containing tha sites for attachment of tha beads. This may be done in a variety of buffers, including equeous end orgenic solvents, and mixtures. The solvent can evaporete, and excess beads are removed.

In a preferred embodiment, when non-covalent methods ara used to associate the beads with the array, e novel method of loading the beads onto the erray is used. This method comprises exposing

the array to e solution of particles (including microspheres and cells) and then epplying energy, e.g. agitating or vibrating the mixture. This rasults in an array comprising more tightly associated particias, as the agitation is dona with aufficiant anergy to cause weakly-associated beads to fall off (or out, in the casa of wells). Thesa sites ara than available to bind a differant baad. In this way, beads that exhibit a high affinity for the sites ere eelected. Arrays mede in this way heve two main edventages as compared to a more static loading: first of all, a higher percentage of tha sites can ba filled easily, and secondly, the arrays thus loaded show a substantial decrease in bead loss during assays. Thus, in a preferrad embodiment, thesa methods ara used to generate arrays that hava at laast about 50% of the sites filled, with at least ebout 75% being preferred, end at least about 90% being particularly praferred. Similarly, arrays ganarated in this manner preferably losa lass than about 20% of tha baads during an assay, with less than about 10% being preferred and less than about 5% being particularly preferrad.

In this embodimant, tha substrata comprising the surface with tha discrata sites is immarsad into a solution comprising the particles (baads, cells, etc.). The surfaca may comprisa wells, as is described harain, or other types of sites on a patterned surfaca such that thare is a differential affinity for the sites. This differetial affinity results in a competitive process, such that particles that will associate mora tightly ara selected. Preferably, the entira surfaca to ba "loaded" with beads is in fluid contact with tha solution. This solution is ganerally a slurry ranging from about 10,000:1 beads:solution (vol:vol) to 1:1. Generally, tha solution can comprisa any number of reagents, including aqueous buffars, organic solvents, salts, other reagent components, etc. In addition, the solution preferebly comprisas an excess of baads; that is, thera ara mora beads than sitas on tha array. Praferred ambodiments utilize two-fold to billion-fold axcass of beads.

The immarsion can mimic the essey conditions; for example, if the erray is to be "dipped" from above into a microtiter plate comprising samplas, this configuration can be repeated for tha loading, thus minimizing the beads that ara likaly to fall out due to gravity.

Once the surface has been immersed, the substrate, the solution, or both ere subjected to e competitive process, wherаby the particles with lowar affinity can ba disassociated from tha substrate and replaced by particias exhibiting a highar affinity to tha sita. This compatitiva process is done by tha introduction of enargy, in the form of heat, sonication, stirring or mixing, vibrating or agitating the solution or substrata, or both.

A praferrad embodimant utilizas agitation or vibration. In ganaral, tha amount of manipulation of the substrata is minimized to pravant damage to tha array; thus, preferred ambodimants utiliza the agitation of the solution rathar than tha array, although aither will work. As will be appreciatad by those in the art, this agitation can take on any number of forms, with a prefarred ambodiment utilizing microtiter plates comprising baad solutions being agitated using microtitar plate shakers.

The agitation proceeds for a period of time sufficient to load the array to a desired fill. Depending on the size end concentration of the beads end the size of the array, this time may range from about 1 second to days, with from about 1 minute to about 24 hours being preferred.

It should be notad that not ell sites of an erray may comprise a baad; that is, thera may be some sites on the substrate surface which are empty. In addition, there may be some sites that contain more than one bead, although this is not preferred.

In soma embodimants, for example when chemical attachment is dona, it is possibla to attach the beads in a non-random or ordered way. For example, using photoactivatibla attachmant linkars or photoactivatibla adhesives or masks, selected sites on tha array may be sequentielly rendered suitabla for attachmant, such that dafined populations of beads are laid down.

The arrays of the present invention are constructed such that information about the identity of the capture probe is built into the array, such thet the random deposition of the beads in the fiber wells can be "decoded" to ellow identificetion of the cepture probe at all positions. This may be done in e variety of ways, and either before, during or after the use of the array to detect target molecules.

Thus, after the erray is made, it is "decoded" in order to identify the locetion of one or more of the capture probes, i.e. each subpopulation of beads, on the substrate surface.

In e preferred embodiment, pyrosequencing techniques ere used to decode the erray, es is generally described in "Nucleic Acid Sequencing Using Microsphere Arrays", filed October 22, 1999 (no U.S.S.N. received yet), hereby expressly incorporated by reference.

In a preferred embodiment, e selective decoding system is used. In this case, only those microspheres exhibiting a chenge in the opticel signel as a result of the binding of a target sequence are decoded. This is commonly done when the number of "hits", i.e. the number of sites to decode, is generally low. Thet is, the erray is first scanned under experimental conditions in the absence of tha target sequences. Tha sample contalning the target sequences is added, and only those locations exhibiting a change in the optical signal are decoded. For example, the beads at either the positive or negative signal locetions may be aither selectively tagged or released from the array (for axample through the use of photocleavable linkers), and subsequentiy sorted or enriched in a fluorescence-activated cell sorter (FACS). Thet is, either all tha negative beads are raleased, and then the positive beads are either relaasad or analyzed in situ, or alternatively all tha positives ara released and analyzed. Alternatively, the labels may comprise halogenated aromatic compounds, and detection of the label is done using for example ges chromatography, chemical tegs, isotopic tags mass spectral tags.

As will be appreciated by those in the art, this may also be done in systems where the erray is not decoded; i.e. there need not ever be e correletion of bead composition with location. In this embodiment, the beads ere loeded on the errey, and the essey is run. The "positives", i.e. those beeds displeying e chenge in the opticel signel es is more fully outlined below, ere then "marked" to distinguish or eeparate them from the "negative" beeds. This can be done in several ways, preferably using fiber optic errays. In e preferred embodiment, each beed contains e fluorescent dye. After the essey and the identification of the "positives" or "active beads", light is shown down either only the positive fibers or only the negetive fibers, generelly in the presence of a light-activeted reegent (typically dissolved oxygen). In the former cese, ell the active beads are photobleached. Thus, upon non-selective release of all the beads with subsequent sorting, for example using e fluorescence activated cell sorter (FACS) machine, the non-fluorescent active beads can be sorted from the fluorescent negetive beads. Alternatively, when light is shown down the negetive fibers, all the negetives ere non-fluorescent end the the postives ere fluorescent, and sorting can proceed. The charecterizetion of the ettached capture probe mey be done directly, for example using mass spectroscopy.

Alternetively, the identificetion may occur through the use of identifier moieties ("IMs"), which ere similar to IBLs but need not necessarily bind to DBLs. That is, rather than elucidete the structure of the capture probe directly, the composition of the IMs may serve as the identifier. Thus, for example, a specific combination of IMs can serve to code the bead, and be usad to identify the egent on the beed upon release from the beed followed by subsequent enalysis, for example using e gas chrometogreph or mass spectroscope.

Altemetively, rather than heving each beed contain a fluorescent dye, each baad comprises a non-fluorescent precursor to e fluorescent dye. For exemple, using photocleavable protecting groups, such as certein ortho-nitrobenzyl groups, on e fluorescent molecule, photoactivation of the fluorochrome cen be done. After the assey, light is shown down again either the "positive" or the "negative" fibers, to distinquish these populations. The illuminated precursors are then chemically converted to e fluorescent dye. All the beads ere then released from the erray, with sorting, to form populetions of fluorescent end non-fluorescent beads (either the positives and the negatives or vice versa).

In an elternate preferred embodiment, the sites of attachmant of the beads (for exemple the wells) include e photopolymerizable reagent, or the photopolymerizeble agent is edded to the assembled erray. Afler the test essay is run, light is shown down egain either the "positive" or the "negetive" fibers, to distinquish these populetions. As a result of the irradiation, either ell the positives or ell the negatives ere polymerized end trapped or bound to the sites, while the other population of beads can be releesed from the array.

In e preferred embodiment, the location of every cepture probe is determined using decoder binding ligends (DBLs). As outlined above, DBLs ere binding ligends that will either bind to identifier binding ligends, if present, or to the capture probes themselves, preferably when the capture probe is e nucleic ecid or protein.

In e preferred embodiment, es outlined above, the DBL binds to the IBL.

In a preferred embodiment, the cepture probes ere single-strended nucleic ecids end the DBL is e substantially complementary single-strended nucleic acid thet binds (hybridizes) to the capture probe, termed e decoder probe herein. A decoder probe thet is substantially complementary to eech candidete probe is mede end used to decode the erray. In this embodiment, the candidate probes end the decoder probes should be of sufficient length (and the decoding step run under suitable conditions) to ellow specificity; i.e. each candidate probe binds to its corresponding decoder probe with sufficient specificity to allow the distinction of each candidete probe.

In e preferred embodiment, the DBLs ere either directly or indirectly labeled. In e preferred embodiment, the DBL is directly labeled, that is, the DBL comprises e lebel. In an eltemete embodiment, the DBL is indirectly lebeled; thet is, e labeling binding ligend (LBL) thet will bind to the DBL is used. In this embodiment, the lebeling binding ligand-DBL pair can be as described ebove for IBL-DBL pairs.

Accordingly, the identification of the locetion of the individual beads (or subpopulations of beeds) is done using one or more decoding steps comprising a binding between the labeled DBL end either the IBL or the capture probe (i.e. e hybridizetion between the candidate probe and the decoder probe when the capture probe is e nucleic ecid). After decoding, the DBLs can be removed end the erray can be used; however, in some circumstances, for exemple when the DBL binds to en IBL end not to the capture probe, the removal of the DBL is not required (elthough it may be desireble in some circumstences). In addition, as outlined herein, decoding mey be done either before the erray is used to in en essey, during the essey, or efter the essay.

In one embodiment, e single decoding step is done. In this embodiment, eech DBL is labeled with e unique label, such thet the the number of unique tags is equal to or greater than the number of capture probes (although in some ceses, "reuse" of the unique labels cen be done, as described herein; similarly, minor variants of candidate probes can shere the seme decoder, if the variants ere encoded in enother dimension, i.e. in the bead size or lebel). For eech capture probe or IBL, a DBL is made thet will specifically bind to it and contains e unique tag, for example one or more fluorochromes. Thus, the identity of each DBL, both its composition (i.e. its sequence when it is a nucleic acid) and its lebel, is known. Then, by edding the DBLs to the errey containing the capture probes under conditions which allow the formation of complexes (termed hybridization complexes when the components ere

nucleic acids) between the DBLs and either the capture probes or the IBLs, the location of each DBL can be elucidated. This allows the identification of the location of each capture probe; the random array has been decoded. The DBLs can then be removed, if necessary, and the target sample applied.

In a preferred embodiment, the number of unique labels is less than the number of unique capture probes, and thus a sequential series of decoding steps are used. In this embodiment, decoder probes are divided into n sets for decoding. The number of sets corresponds to the number of unique tags. Each decoder probe is labeled in n separate reactions with n distinct tags. All the decoder probes share the same n tags. The decoder probes are pooled so that each pool contains only one of the n tag versions of each decoder, and no two decoder probes have the same sequence of tags across all the pools. The number of pools required for this to be true is determined by the number of decoder probes and the n. Hybridization of each pool to the array generates a signal at every address. The sequential hybridization of each pool in turn will generate a unique, sequence-specific code for each candidate probe. This identifies the candidate probe at each address in the array. For example, if four tags are used, then 4 X n sequential hybridizations can ideally distinguish 4^n sequences, although in some cases more steps may be required. After the hybridization of each pool, the hybrids are denatured and the decoder probes removed, so that the probes are rendered single-stranded for the next hybridization (although it is also possible to hybridize limiting amounts of target so that the available probe is not saturated. Sequential hybridizations can be carried out and analyzed by subtracting pre-existing signal from the previous hybridization).

An example is illustrative. Assuming an array of 16 probe nucleic acids (numbers 1-16), and four unique tags (four different fluors, for example; labels A-D). Decoder probes 1-16 are made that correspond to the probes on the beads. The first step is to label decoder probes 1-4 with tag A, decoder probes 5-8 with tag B, decoder probes 9-12 with tag C, and decoder probes 13-16 with tag D. The probes are mixed and the pool is contacted with the array containing the beads with the attached candidate probes. The location of each tag (and thus each decoder and candidate probe pair) is then determined. The first set of decoder probes are then removed. A second set is added, but this time, decoder probes 1, 5, 9 and 13 are labeled with tag A, decoder probes 2, 6, 10 and 14 are labeled with tag B, decoder probes 3, 7, 11 and 15 are labeled with tag C, and decoder probes 4, 8, 12 and 16 are labeled with tag D. Thus, those beads that contained tag A in both decoding steps contain candidate probe 1; tag A in the first decoding step and tag B in the second decoding step contain candidate probe 2; tag A in the first decoding step and tag C in the second step contain candidate probe 3; etc. In one embodiment, the decoder probes are labeled in situ; that is, they need not be labeled prior to the decoding reaction. In this embodiment, the incoming decoder probe is shorter than the candidate probe, creating a 5' "overhang" on the decoding probe. The addition of labeled ddNTPs (each labeled with a unique tag) and a polymerase will allow the addition of the tags in a sequence specific manner, thus creating a sequence-specific pattern of signals. Similarly, other modifications can be done,

including ligation, etc.

In addition, since the size of the array will be set by the number of unique decoding binding ligands, it is possible to "reuse" a set of unique DBLs to allow for a greater number of test sites. This may be done in several ways; for example, by using some subpopulations that comprise optical signatures. Similarly, the use of a positional coding schema within an array; different sub-bundles may reuse the set of DBLs. Similarly, one embodiment utilizes bead size as a coding modality, thus allowing the reuse of the set of unique DBLs for each bead size. Alternatively, sequential partial loading of arrays with beads can also allow the reuse of DBLs. Furthermore, "code sharing" can occur as well.

In a preferred embodiment, the DBLs may be reused by having some subpopulations of beads comprise optical signatures. In a preferred embodiment, the optical signature is generally a mixture of reporter dyes, preferably fluorescent. By varying both the composition of the mixture (i.e. the ratio of one dye to another) and the concentration of the dye (leading to differences in signal intensity), matrices of unique optical signatures may be generated. This may be done by covalently attaching the dyes to the surface of the beads, or alternatively, by entrapping the dye within the bead.

In a preferred embodiment, the encoding can be accomplished in a ratio of at least two dyes, although more encoding dimensions may be added in the size of the beads, for example. In addition, the labels are distinguishable from one another; thus two different labels may comprise different molecules (i.e. two different fluors) or, alternatively, one label at two different concentrations or intensity.

In a preferred embodiment, the dyes are covalently attached to the surface of the beads. This may be done as is generally outlined for the attachment of the capture probes, using functional groups on the surface of the beads. As will be appreciated by those in the art, these attachments are done to minimize the effect on the dye.

In a preferred embodiment, the dyes are non-covalently associated with the beads, generally by entrapping the dyes in the pores of the beads.

Additionally, encoding in the ratios of the two or more dyes, rather than single dye concentrations, is preferred since it provides insensitivity to the intensity of light used to interrogate the reporter dye's signature and detector sensitivity.

In a preferred embodiment, a spatial or positional coding system is done. In this embodiment, there are sub-bundles or subarrays (i.e. portions of the total array) that are utilized. By analogy with the telephone system, each subarrey is an "area code", that can have the same tags (i.e. telephone numbers) of other subarrays, that are separated by virtue of the location of the subarrey. Thus, for example, the same unique tags can be reused from bundle to bundle. Thus, the use of 50 unique tags

in combination with 100 different subarrays can form an array of 5000 differant capture probes. In this embodiment, it becomes important to be able to identify one bundle from another; in generel, this is done either menually or through the use of marker beads, i.e. beeds contalning unique tags for aach subarray.

In altarnative embodiments, additional encoding perameters can be added, such as microsphere size. For example, the use of different size beads may also ellow the reuse of sets of DBLs; that is, it is possible to use microspheres of different sizes to expand tha encoding dimensions of the microspheres. Optical fiber arrays can be febricated contalning pixels with different fiber diemeters or cross-sections; altemetively, two or more fiber optic bundles, each with different cross-sections of the individual fibars, can be edded together to form a larger bundle; or, fiber optic bundles with fiber of the same size cross-sections can be used, but just with different sized beads. With different diameters, tha largest wells can be filled with the largest microspheres and then moving onto progressively smaller microspheres in the smaller walls until all size wells are then filled. In this manner, the same dye ratio could be used to encode microspheres of diffarent sizes thereby axpanding the number of different oligonucleotide sequences or chemical functionalities present in the array. Although outlined for fiber optic substrates, this as well as the other methods outlined herein can be used with other substretes and with other attachment modalities as well.

In a preferred embodiment, the coding and decoding is accomplished by sequential loading of the microspheres into the erray. As outlined ebove for spatial coding, in this embodimant, the optical signetures can be "reused". In this embodiment, the library of microspheres each comprising a different capture proba (or the subpopulations each comprise a different capture probe), is divided into a plurality of sublibraries; for example, depending on the siza of the desired array and the number of uniqua tags, 10 sublibraries each comprising roughly 10% of the total library may be made, with each sublibrary comprising roughly the same unique tags. Then, the first sublibrary is added to the fiber optic bundle comprising the wells, and the location of each capture probe is determined, generally through the use of DBLs. The second sublibrary is then edded, and the location of each capture probe is again determined. The signal in this case will comprise tha signal from tha "first" DBL and the "second" DBL; by comparing the two matrices tha location of each bead in each sublibrary can be determined. Similarly, adding the third, fourth, etc. sublibraries sequentially will allow the array to be filled.

In a preferred embodiment, codes cen be "shared" in several ways. In a first embodiment, a singla code (i.e. IBL/DBL pair) can be assigned to two or more agents if the target sequences different sufficiently in their binding strengths. For exampie, two nucleic ecid probes used in en mRNA quantitation assay can share the same code if the ranges of their hybridization signel intensities do not overlap. This can occur, for example, when one of tha target sequences is always present at a much higher concentration than the other. Alternatively, the two target sequences might elways be present

at a similar concentration, but differ in hybridization efficiency.

Alternatively, a single code can be assigned to multiple agents if the agents are functionally equivalent. For example, if a set of oligonucleotide probes are designed with the common purpose of detecting the presence of a particular gene, then the probes are functionally equivalent, even though they may differ in sequence. Similarly, an array of this type could be used to detect homologs of known genes. In this embodiment, each gene is represented by a heterologous set of probes, hybridizing to different regions of the gene (and therefore differing in sequence). The set of probes share a common code. If a homolog is present, it might hybridize to some but not all of the probes. The level of homology might be indicated by the fraction of probes hybridizing, as well as the average hybridization intensity. Similarly, multiple antibodies to the same protein could all share the same code.

In a preferred embodiment, decoding of self-assembled random arrays is done on the bases of pH titration. In this embodiment, in addition to capture probes, the beads comprise optical signatures, wherein the optical signatures are generated by the use of pH-responsive dyes (sometimes referred to herein as "ph dyes") such as fluorophores. This embodiment is similar to that outlined in PCT US98/05025 and U.S.S.N. 09/151,877, both of which are expressly incorporated by reference, except that the dyes used in the present invention exhibits changes in fluorescence intensity (or other properties) when the solution pH is adjusted from below the pKa to above the pKa (or vice versa). In a preferred embodiment, a set of pH dyes are used, each with a different pKa, preferably separated by at least 0.5 pH units. Preferred embodiments utilize a pH dye set of pKa's of 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11, and 11.5. Each bead can contain any subset of the pH dyes, and in this way a unique code for the capture probe is generated. Thus, the decoding of an array is achieved by titrating the array from pH 1 to pH 13, and measuring the fluorescence signal from each bead as a function of solution pH.

Thus, the present invention provides array compositions comprising a substrate with a surface comprising discrete sites. A population of microspheres is distributed on the sites, and the population comprises at least a first and a second subpopulation. Each subpopulation comprises a capture probe, and, in addition, at least one optical dye with a given pKa. The pKas of the different optical dyes are different.

In a preferred embodiment, "random" decoding probes can be made. By sequential hybridizations or the use of multiple labels, as is outlined above, a unique hybridization pattern can be generated for each sensor element. This allows all the beads representing a given clone to be identified as belonging to the same group. In general, this is done by using random or partially degenerate decoding probes, that bind in a sequence-dependent but not highly sequence-specific manner. The process can be repeated a number of times, each time using a different labeling entity, to generate a different pattern of signals based on quasi-specific interactions. In this way, a unique optical signature

is eventuelly built up for each sensor element. By epplying pattern recognition or clustering elgorithms to the optical signetures, the beads can be grouped into sets thet share the seme signeture (i.e. cerry the same probes).

In order to identify the actual sequence of the clone itself, edditionel procedures are required; for exemple, direct sequencing cen be done, or an ordered erray contalning the clones, such es a spotted cDNA erray, to generete e "key" that links e hybridizetion pattern to a specific clone.

Altemetively, clona arrays can be decoded using binery decoding with vector tags. For example, partielly randomized oligos ere cloned into e nucleic ecid vector (e.g. plasmid, phage, etc.). Eech oligonucleotide sequence consists of e subset of a limited set of sequences. For exemple, if the limites set comprises 10 sequences, each oligonucleotide mey heve some subset (or all of the 10) sequences. Thus eech of the 10 sequences can be presant or absent in the oligonucleotide. Therefore, there are 2^{10} or 1,024 possible combinations. The sequences may overlap, and minor variants can also be represented (e.g. A, C, T end G substitutions) to increase the number of possible combinetions. A nucleic acid librery is cloned into a vector containing the random code sequences. Altematively, other methods such as PCR can be used to add the tags. In this way it is possible to use a small number of oligo decoding probes to decode en array of clones.

As will be appreciated by those in the art, the systems of the invention mey take on a large number of different configurations, as is generally depicted in the Figures. In general, there are three types of systems thet can be used: (1) "non-sandwich" systems (also referred to herein as "direct" detection) in which the target sequence itself is labeled with detectable labels (again, either beceuse the primers comprise labels or due to the incorporation of labels into the newly synthesized strand); (2) systems in which label probes directly bind to the target enelytes; and (3) systems in which label probes ere indirectly bound to the target sequences, for exemple through the use of emplifier probes.

Detection of the reactions of the invention, including the direct detection of products end indirect detection utilizing lebel probes (i.e. sendwich essays), is praferably done by detecting assay complexes comprising detectable lebels, which cen be ettached to the essey complex in a variety of ways, as is more fully described below.

Once the target sequence has preferably been anchored to the array, an emplifier probe is hybridized to the target sequence, either directly, or through the use of one or more label extender probes, which serves to ellow "generic" amplifier probes to be made. As for ell the steps outlined herein, this mey be done simultaneously with capturing, or sequentielly. Preferably, the emplifier probe contains a multiplicity of amplification sequences, elthough in some embodiments, es described below, the emplifier probe may contain only a single emplification sequence, or at least two emplification sequences. The emplifier probe may teke on a number of different forms; either a branched

conformation, a dendrimer conformation, or a linear "string" of amplification sequences. Label probes comprising detectabla labels (preferably but not required to be fluorophores) then hybridize to the amplification sequences (or in some cases the label probes hybridize directly to the target sequence), end the labels detected, aa is more fully outlined below.

Accordingly, the present invention provides compositions comprising an amplifier probe. By "amplifier probe" or "nucleic acid multimer" or "amplification multimer" or grammatical equivalents herein is meant e nucleic ecid probe that is used to facilitate signal amplification. Amplifier probes comprise at least e first single-stranded nucleic acid probe sequence, es defined below, and at least one single-stranded nucleic acid amplification sequence, with e multiplicity of amplification sequences being preferred.

Amplifier probes comprise a first probe sequence thet is used, either directly or indirectly, to hybridize to the target sequence. That is, the amplifier probe itself may have a first probe sequence that is substantially complementary to the target sequence, or it has a first probe sequence that is substantially complementary to a portion of an edditional probe, in this case called a label extender probe, that has a first portion that is substantially complementary to the target sequence. In a preferred embodiment, the first probe sequence of the amplifier probe is substantially complementary to the target sequence.

In general, as for all the probes herein, the first probe sequence is of a length sufficient to give specificity and stability. Thus generally, the probe sequences of the invention that are designed to hybridize to another nucleic acid (i.e. probe sequences, amplification sequences, portions or domains of larger probes) are at least about 5 nucleosides long, with at least ebout 10 being preferred and at leest about 15 being especially preferred.

In a preferred embodiment, several different emplifier probes ere used, each with first probe sequences thet will hybridize to a different portion of the target sequence. That is, there is more than one level of amplification; the amplifier probe provides en emplification of signal due to a multiplicity of labelling events, end several different amplifier probes, each with this multiplicity of labels, for each target sequence is used. Thus, preferred embodiments utilize at least two different pools of amplifier probes, each pool having e different probe sequence for hybridization to different portions of the target sequence; the only reel limitation on the number of different emplifier probes will be the length of the original terget sequence. In addition, it is elso possible that the different amplifier probes contain different amplification sequences, although this is generally not preferred.

In e preferred embodiment, the amplifier probe does not hybridize to the sample terget sequence directly, but instead hybridizes to e first portion of e label extender probe. This is particularly useful to allow the use of "genaric" emplifier probes, that is, amplifier probes that can be used with a variety of

different targets. This may be desireble since several of the emplifier probee require special synthesis techniques. Thus, the eddition of e relatively short probe es e label extender probe is preferred. Thus, the first probe sequence of the emplifier probe is substantially complementary to e first portion or domain of a first label extender single-stranded nucleic ecid probe. The label extender probe elso contains e second portion or domein that is substantielly complementary to e portion of the target sequence. Both of these portions ere preferebly et leest ebout 10 to ebout 50 nucleotides in length, with a range of ebout 15 to about 30 being preferred. The terms "first" and "second" ere not meent to confer an orientation of the sequences with respect to the 5'-3' orientation of the target or probe sequences. For example, essuming e 5'-3' orientation of the complementary target sequence, the first portion mey be located either 5' to the second portion, or 3' to the second portion. For convenience herein, the order of probe sequences ere generally shown from left to right.

In e preferred embodiment, more then one label extender probe-emplifier probe peir may be used, that is, n is more then 1. That is, e plurality of lebel extender probes may be used, each with a portion that is substantielly complementary to a different portion of the target sequence; this can serve as enother level of amplificetion. Thus, e preferred embodiment utilizes pools of at least two label extender probes, with the upper limit being set by the length of the target sequence.

In a preferred embodiment, more then one label extender probe is used with a single emplifier probe to reduce non-epecific binding, as is generally outlined in U.S. Patent No. 5,681,697, incorporeted by reference herein. In this embodiment, a first portion of the first lebel extender probe hybridizes to e first portion of the target sequence, end the second portion of the first lebel extender probe hybridizes to a first probe sequence of the amplifier probe. A first portion of the second label extender probe hybridizes to e second portion of the target sequence, and the second portion of the second lebel extender probe hybridizes to a second probe sequence of the emplifier probe. These form structures sometimes referred to es "cruciform" structures or configurations, end are generally done to confer stability when large branched or dendrimeric amplifier probes are used.

In eddition, es will be eppreciated by those in the ert, the label extender probes may interact with a preamplifier probe, described below, rather than the amplifier probe directly.

Simllerly, as outlined above, a preferred embodiment utilizes several different emplifier probes, each with first probe sequences that will hybridize to e different portion of the label extender probe. In addition, as outlined ebove, it is also possible that the different emplifier probes contain different emplification sequences, elthough this is generally not preferred.

In eddition to the first probe sequence, the emplifier probe elso comprises et leest one amplification sequence. An "amplification sequence" or "emplificetion segment" or grammatical equivalents herein is meant e sequence thet is used, either directly or indirectly, to bind to a first portion of a label probe

as is more fully described below. Preferably, the amplifier probe comprises a multiplicity of amplification sequences, with from about 3 to about 1000 being preferred, from about 10 to about 100 being particularly preferred, and about 50 being especially preferred. In some cases, for example when linear amplifier probes are used, from 1 to about 20 is preferred with from about 5 to about 10 being particularly preferred.

The amplification sequences may be linked to each other in a variety of ways, as will be appreciated by those in the art. They may be covalently linked directly to each other, or to intervening sequences or chemical moieties, through nucleic acid linkages such as phosphodiester bonds, PNA bonds, etc., or through interposed linking agents such amino acid, carbohydrate or polyol bridges, or through other cross-linking agents or binding partners. The site(s) of linkage may be at the ends of a segment, and/or at one or more internal nucleotides in the strand. In a preferred embodiment, the amplification sequences are attached via nucleic acid linkages.

In a preferred embodiment, branched amplifier probes are used, as are generally described in U.S. Patent No. 5,124,246, hereby incorporated by reference. Branched amplifier probes may take on "fork-like" or "comb-like" conformations. "Fork-like" branched amplifier probes generally have three or more oligonucleotide segments emanating from a point of origin to form a branched structure. The point of origin may be another nucleotide segment or a multifunctional molecule to which at least three segments can be covalently or tightly bound. "Comb-like" branched amplifier probes have a linear backbone with a multiplicity of sidechain oligonucleotides extending from the backbone. In either conformation, the pendant segments will normally depend from a modified nucleotide or other organic moiety having the appropriate functional groups for attachment of oligonucleotides. Furthermore, in either conformation, a large number of amplification sequences are available for binding, either directly or indirectly, to detection probes. In general, these structures are made as is known in the art, using modified multifunctional nucleotides, as is described in U.S. Patent Nos. 5,635,352 and 5,124,246, among others.

In a preferred embodiment, dendrimer amplifier probes are used, as are generally described in U.S. Patent No. 5,175,270, hereby expressly incorporated by reference. Dendrimeric amplifier probes have amplification sequences that are attached via hybridization, and thus have portions of double-stranded nucleic acid as a component of their structure. The outer surface of the dendrimer amplifier probe has a multiplicity of amplification sequences.

In a preferred embodiment, linear amplifier probes are used, that have individual amplification sequences linked end-to-end either directly or with short intervening sequences to form a polymer. As with the other amplifier configurations, there may be additional sequences or moieties between the amplification sequences. In one embodiment, the linear amplifier probe has a single amplification sequence.

In addition, the amplifier probe may be totally linear, totally branched, totally dendrimeric, or any combination thereof.

The amplification sequences of the amplifier probe are used, either directly or indirectly, to bind to a label probe to allow detection. In a preferred embodiment, the amplification sequences of the amplifier probe are substantially complementary to e first portion of e label probe. Alternatively, emplifier extender probes ara used, thet have e first portion that binds to the amplification sequence and a second portion that binds to the first portion of the label probe.

In addition, the compositions of the Invention may include "preamplifier" molecules, which eerves a bridging molety between the label extender molecules and the amplifier probes. In this way, mora amplifier and thus more lebels are ultimately bound to the detection probes. Preamplifier molecules mey be either linear or branched, and typically contain in the range of about 30-3000 nucleotides.

Thus, label probes era either substantielly complementary to en amplification sequence or to a portion of the target sequence.

Detection of the nucleic acid reactions of the Invention, including the direct detection of genotyping products and Indirect detection utilizing label probes (i.e. sandwich asseys), is done by detacting assay complexes comprising labels.

In a preferred embodimant, several levels of redundancy are built Into the arrays of the invention. Building redundancy Into an erray gives several significant advantages, including tha ability to meke quantitative estimetes of confidence about the data and signficant Increases In sensitivity. Thus, preferred embodiments utilize erray redundancy. As will be appreciated by those in the art, there are at least two types of radundancy thet cen be built into an errey: the use of multiple identicel sensor elements (termed herein "sensor redundancy"), and the use of multiple sensor elements directed to the sama target enalyte, but comprising different chemical functionalities (termed herein "target redundancy"). For example, for the detection of nucleic acids, sensor redundancy utilizes of a plurality of sensor elements such es beeds comprising Identical binding ligands such as probes. Target redundancy utilizes eensor elements with different probes to the same target: one probe may span the first 25 bases of the target, e second probe may span the second 25 bases of the target, etc. By building in either or both of these types of redundancy Into en array, significant benefits are obtained. For example, a variety of statistical mathematical analyses may be done.

In eddition, whila this is generally described herein for bead arrays, as will be appreciated by those in the art, this techniques can be used for any type of errays designed to detect target analytas. Furthermore, while these techniques are generally described for nucleic ecid systems, these techniques ere useful in the detection of other binding ligand/target analyte systems es well.

In e preferred embodiment, sensor redundency is used. In this embodiment, a plurality of sensor elements, e.g. beads, comprising identical bioactive egents ere usad. That is, eech subpopuletion comprises a plurelity of beeds comprising identical bioactive egents (e.g. binding ligends). By using e number of identical aensor elements for e given array, the optical signel from each sensor element cen be combined and any number of statistical analyses run, as outlined below. This can be done for a variety of reesons. For example, in time varying meesuremants, redundancy can significently reduce the noise in the system. For non-time based measurements, redundency can significantly increese the confidence of the data.

In e preferred embodiment, a plurelity of identicel sensor elements are used. As will be eppreciated by those in the ert, the number of identical sensor elements will vary with the epplication end use of the sensor array. In general, enywhere from 2 to thousands may be used, with from 2 to 100 being preferred, 2 to 50 being particularly preferred and from 5 to 20 being especially preferred. In general, preliminary results indicate thet roughly 10 beads gives e sufficient edvantage, elthough for some applications, more identical sensor elements can be used.

Once obtained, the opticel response signels from e plurality of sensor beads within each bead subpopuletion cen be menipulated end enelyzed in a wide variety of ways, including baseline edjustment, everaging, standard deviation enelysis, distribution and cluster enalysis, confidence interval enalysis, meen testing, etc.

In a prafarred embodiment, the first manipulation of the optical response signals is an optional baseline edjustment. In a typical procedure, the standardized optical responses ere edjusted to start et e value of 0.0 by subtrecting the integer 1.0 from ell data points. Doing this allows the baselne-loop data to remain et zero even when summed together end the random response signel noise is canceled out. When the semple is a fluid, the fluid pulse-loop temporal region, however, frequently exhibits e characteristic chenge in response, either positive, negative or neutral, prior to the sample pulse and often requires a baseline edjustment to overcome noise associeted with drift in the first few data points due to charge buildup in tha CCD camera. If no drift is present, typically the baseline from the first date point for eech beed sensor is subtracted from all the response data for the same bead. If drift is observed, the average baseline from the first ten deta points for each bead sensor is substracted from the ell the response deta for the same baed. By applying this baseline adjustmant, when multiple beed responses ere added together they can be emplified while the baselne remains at zero. Since all beads respond et the sema time to the semple (e.g. the sample pulse), they all see the pulse at the axact same time end there is no registering or adjusting needed for overlaying their responses. In eddition, other types of baseline edjustment mey be done, depanding on the requirements end output of the system used.

Once the beseline hae been edjusted, a number of possible statistical analyses may be run to generate known statistical parameters. Analyses besed on redundancy are known and generally dascribed in texts such es Freund and Welpole, Mathematical Statistics, Prentica Hall, Inc. New Jersey, 1980, hereby incorporated by raference in its entirety.

In a preferred embodiment, signal summing is done by simply adding the intensity values of all responses at aach tima point, generating a new temporal responsa comprised of the sum of all bead responses. These values can be baseline-edjusted or raw. As for all the analyses described herein, signal summing can ba performed in raal time or during post-data acquisition data reduction and analysis. In one embodiment, signal summing is performed with a commercial spreadsheet program (Excel, Microsoft, Redmond, WA) efter optical responsa data is collactad.

In e preferred embodiment, cummulative response deta is generated by simply adding all data points in successive time intervals. This final column, comprised of tha sum of all data points at e particular time interval, mey than be compared or plotted with the individual bead responses to determine the extent of signal enhancemant or improved signal-to-noise ratios.

In a preferred embodiment, the mean of the subpopulation (i.e. the plurality of identical beads) is determined, using the well known Equation 1:

Equation 1

$$\mu = \sum \frac{x_i}{n}$$

In some embodiments, the subpopulation mey be redefined to exclude some beads if necessary (for example for obvious outliars, as discussed balow).

In a praferred ambodiment, the standard deviation of the subpopulation can be determined, genarally using Equation 2 (for the entire subpopulation) and Equation 3 (for less than the entire subpopuletion):

Equation 2

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Equation 3

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

As for the meen, the subpopuletion mey be redefined to exclude some beeds if necessary (for example for obvious outliers, as discussed below).

In e praferred embodiment, statistical analyses are done to evaluate whether a particular data point has statistical validity within e subpopulation by using techniques including, but not limitad to, t distribution and cluster analysis. This may be done to statistically discard outliers that mey otherwise skew the result and increase the signal-to-noise ratio of any particular experimant. This may be done using Equation 4:

Equation 4

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

In e preferred embodiment, the quality of the data is evaluated using confidenca intervals, as is known in the art. Confidence Intervals can be used to facilitate more comprehensive data processing to measura the statistical validity of e result.

In a preferred ambodiment, statistical parameters of a subpopulation of beeds ere used to do hypothesis testing. One application is tests concerning means, also called mean testing. In this application, statisticel evaluation is done to determine whether two subpopulations are different. For example, one sampie could be compered with another sample for each subpopuletion within an arrey to determine if the variation is statistically significant.

In eddition, mean testing can also be used to differentiate two different assays that shara the same code. If the two essays give results thet ere statistically distinct from each other, then the subpopulations that share e common code cen be distinguished from aach other on the basis of the assay and the mean test, shown below in Equation 5:

Equation 5

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Furtharmore, analyzing the distribution of individual members of a subpopulation of sensor elements may ba done. For example, e subpopulation distribution can be evalueted to determine whether the distribution is binomial, Poisson, hypergeometric, etc.

In addition to the sensor redundancy, a preferred embodiment utilizes e plurelity of sensor elements that ere directed to a single target analyte but yet are not identical. For example, a single target nucleic acid analyte may have two or more sensor elements each comprising a different probe. This edds e level of confidence es non-specific binding interactions can be statisticelly minimized. When nucleic acid target analytes ere to be evaluated, the redundant nucleic acid probes may be overlapping, edjacent, or spatially separated. However, it is preferred that two probes do not compete for e single binding site, so adjacent or seperated probes ere preferred. Similarly, when proteineceous target enalytes are to be eveluated, preferred embodiments utilize bioactive agent binding agents that bind to different parts of the target. For example, when antibodies (or antibody fragments) ere used as bioactive agents for the binding of terget proteins, preferred embodiments utilize antibodies to different epitopes.

In this embodiment, a plurality of different sensor elements mey be used, with from ebout 2 to about 20 being preferred, and from about 2 to about 10 being especially preferred, and from 2 to about 5 being particularly preferred, including 2, 3, 4 or 5. Howeve, as above, more may elso be used, depending on the application.

As above, eny number of statistical analyses may be run on the data from target redundant sensors.

One benefit of the sensor element summing (referred to herein as "bead summing" when beads ere used), is the increase in sensitivity that can occur.

In eddition, the present invention is directed to the use of adapter sequences to essemble arrays comprising target analytes. including non-nucleic acid target anelytes. By "target analyte" or "analyte" or grammatical equivalents herein is meant eny molecule, compound or particle to be detected. As outlined below, target analytes preferably bind to binding ligands, as is more fully described below. As will be eppreciated by those in the art, a large number of enalytes may be detected using the present methods; basically, any target enalyte for which a binding ligand, described below, may be made may be detected using the methods of the invention.

Suitable enalytes include orgenic and inorgenic molecules, including biomolecules. In a preferred embodiment, the analyte may be an environmental pollutant (including pesticides, insecticides, toxins, etc.); a chemical (including solvents, polymers, organic materiels, etc.); therapeutic molecules (including therapeutic end ebused drugs, antibiotics, etc.); biomolecules (including hormones, cytokines, proteins, lipids, cerbohydrates, cellular membrane antigens and receptors (neural, hormonal, nutrient, and cell surface receptors) or their ligends, etc); whole cells (including procaryotic (such as pathogenic bacteria) end eukaryotic cells, including mammalian tumor cells); viruses (including retroviruses, herpesviruses, edenoviruses, lentiviruses, etc.); and spores; etc. Particularly preferred analytes ere environmental pollutants; nucleic acids; proteins (including enzymes,

entibodies, entigens, growth factors, cytokines, etc); therapeutic end ebused drugs; cells; end viruses.

In e preferred embodiment, the target enalyte is a protein. As will be epprecieted by those in the art, there ere a large number of possible proteinaceous target analytes thet may be detected using the present invention. By "proteins" or grammaticel equivalents herein is meent proteins, oligopeptides end peptides, derivatives end analogs, including proteins containing non-neturelly occuring emino ecids and emino acid enelogs, end peptidomimetic structures. The side chains mey be in either the (R) or the (S) configuration. In a preferred embodiment, the amino acids are in the (S) or L-configuration. As discussed below, when the protein is used es e binding ligend, it mey be desireble to utilize protein enelogs to retard degradation by sample contaminants.

Suitable protein target analytes include, but are not limited to, (1) immunoglobulins, particulerly IgEs, IgGs end IgMs, end particularly therapeuticelly or diagnostically relevant entibodies, including but not limited to, for example, entibodies to humen elbumin, epolipoproteins (including epolipoprotein E), human chorionic gonedotropin, cortisol, α-fetoprotein, thyroxin, thyroid stimuleting hormone (TSH), entithrombin, entibodies to pharmaceuticals (including entieptileptic drugs (phenytoin, primidone, cerbeniezepin, ethosuximide, valproic ecid, end phenoberbitol), cardioective drugs (digoxin, lidocaine, procainamide, end disopyramide), bronchodilators (theophylline), entibiotics (chloremphenicol, sulfonamides), antidepressents, immunosuppresants, ebused drugs (amphetamine, methemphetamine, cennabinoids, cocaine and opietes) end antibodies to eny number of viruses (including orthomyxoviruses, (e.g. influenza virus), peremyxoviruses (e.g respiratory syncytial virus, mumps virus, measles virus), adenoviruses, rhinoviruses, coronaviruses, reoviruses, togaviruses (e.g. rubella virus), parvoviruses, poxviruses (e.g. variola virus, vaccinia virus), enteroviruses (e.g. poliovirus, coxseckievirus), hepatitis viruses (including A, B end C), herpesviruses (e.g. Herpes simplex virus, varicelle-zoster virus, cytomegalovirus, Epstein-Berr virus), rotaviruses, Norwalk viruses, hentavirus, arenavirus, rhebdovirus (e.g. rabies virus), retroviruses (including HIV, HTLV-I end -II), papovaviruses (e.g. pepilomevirus), polyomaviruses, end picorneviruses, end the like), end becterie (including e wide variety of pathogenic end non-pathogenic prokaryotes of interest including Bacillus; Vibrio, e.g. *V. cholerae;* Escherichia, e.g. Enterotoxigenic *E. coli*, Shigella, e.g. *S. dysenteriae;* Salmonelle, e.g. *S. typhi;* Mycobacterium e.g. *M. tuberculosis, M. leprae*; Clostridium, e.g. *C. botulinum, C. teteni, C. difficile, C.perfringens*; Comyebacterium, e.g. *C. diphtheriae*; Streptococcus, *S. pyogenes, S. pneumoniae;* Staphylococcus, *e.g. S.* aureus; Haemophilus, e.g. *H. influenzae*; Neisseria, e.g. *N. meningitidis, N. gonorrhoeae*; Yersinia, e.g. *G. lambliaY. pestis*, Pseudomonas, e.g. *P.* a*eruginose, P. putide*; Chlemydie, e.g. *C. trachomatis*; Bordetella, e.g. *B. pertussis;* Treponema, e.g. *T. pelladium;* and the like); (2) enzymes (and other proteins), including but not limited to, enzymes used es indicetors of or treetment for heart diseese, including creetine kinese, lactate dehydrogenase, espartate amino transferese, troponin T, myoglobin, fibrinogen, cholesterol, triglycerides, thrombin, tissue plasminogen activator (tPA); pencreatic disease indicators including amylese, lipase, chymotrypsin and trypsin; liver function enzymes and proteins including cholinesterase, bilirubin, end

alkaline phosphotase; aldolase, prostatic acid phosphatase, terminal deoxynucleotidyl transferase, and bacterial and viral enzymes such as HIV protease; (3) hormones and cytokines (many of which serve as ligands for cellular receptors) such as erythropoietin (EPO), thrombopoietin (TPO), the interleukins (including IL-1 through IL-17), insulin, insulin-like growth factors (including IGF-1 and -2), epidermal growth factor (EGF), transforming growth factors (including TGF-α and TGF-β), human growth hormone, transferrin, epidermal growth factor (EGF), low density lipoprotein, high density lipoprotein, leptin, VEGF, PDGF, ciliary neurotrophic factor, prolactin, adrenocorticotropic hormone (ACTH), calcitonin, human chorionic gonadotropin, cortisol, estradiol, follicle stimulating hormone (FSH), thyroid-stimulating hormone (TSH), leutinzing hormone (LH), progeterone, testosterone, ; and (4) other proteins (including α-fetoprotein, carcinoembryonic antigen CEA.

In addition, any of the biomolecules for which antibodies may be detected may be detected directly as well; that is, detection of virus or bacterial cells, therapeutic and abused drugs, etc., may be done directly.

Suitable target analytes include carbohydrates, including but not limited to, markers for breast cancer (CA15-3, CA 549, CA 27.29), mucin-like carcinoma associated antigen (MCA), ovarian cancer (CA125), pancreatic cancer (DE-PAN-2), and colorectal and pancreatic cancer (CA 19, CA 50, CA242).

The adapter sequences may be chosen as outlined above. These adapter sequences can then be added to the target analytes using a variety of techniques. In general, as described above, non-covalent attachment using binding partner pairs may be done, or covalent attachment using chemical moieties (including linkers).

Once the adapter sequences are associated with the target analyte, including target nucleic acids, the compositions are added to an array. In one embodiment a plurality of hybrid adapter sequence/target analytes are pooled prior to addition to an array. All of the methods and compositions herein are drawn to compositions and methods for detecting the presence of target analytes, particularly nucleic acids, using adapter arrays.

Advantages of using adapters include but are not limited to, for example, the ability to create universal arrays. That is, a single array is utilized with each capture probe designed to hybridize with a specific adapter. The adapters are joined to any number of target analytes, such as nucleic acids, as is described herein. Thus, the same array is used for vastly different target analytes. Furthermore, hybridization of adapters with capture probes results in non-covalent attachment of the target nucleic acid to the microsphere. As such, the target nucleic/adapter hybrid is easily removed, and the microsphere/capture probe can be re-used. In addition, the construction of kits is greatly facilitated by the use of adapters. For example, arrays or microspheres can be prepared that comprise the capture

proba; tha adapters can be packaged along with tha microsphares for attachment to any targat anelyte of Interest. Thus, ona naed only ettach tha adeptar to tha targat anelyta and dispersa on the array for tha construction of an array of targat anelytas.

Onca mada, the compositions of the Invention find use In e numbar of epplications. In a preferrad embodimant, the compositions ere used to probe e sample solution for tha presence or ebsence of a target sequanca, Including tha quantificetion of tha emount of target saquanca prasant.

For SNP enalysis, tha ratio of differant lebals at a particular location on tha array indicates tha homozygosity or heterozygosity of tha target sample, assuming tha same concentration of each readout probe is used. Thus, for example, assuming a first readout proba comprising a first base et the readout position with a first detactabla label and a sacond raadout proba comprising a second base et the readout position with a second detectable label, equal signals (roughly 1:1 (taking into account the different signal intensitias of tha differant labels, different hybridization afficiencies, and other reasons)) of tha first and second labels indicates a heterozygote. The absenca of a signal from tha first labal (or a ratio of epproximately 0:1) indicetes a homozygota of the second datection basa; tha absenca of a signal from the sacond label (or e ratio of approximately 1:0) indicates a homozygote for the first detection base. As is appreciated by those in tha art, tha actual ratios for eny particular system are generally determinad empirically. The ratios also allow for SNP quantitation

The prasent invention also finds use as a methodology for the detection of mutations or mismatches in target nucleic acid sequances. For exampla, recant focus has been on the enalysis of the ralationship betwaan ganetic variation end phenotype by making usa of polymorphic DNA markers. Previous work utilized short tandem repeats (STRs) as polymorphic positional markers; however, recent focus is on the use of singla nucleotida polymorphisms (SNPs), which occur at an averege frequency of more than 1 per kilobase In human genomic DNA. Some SNPs, particularly those in and around coding saquences, are likely to be the diract ceuse of therapeutically relevant phenotypic variants. Thera are a number of well known polymorphisms that cause clinically Important phenotypas; for axampla, the apoE2/3/4 variants ere associated with diffarant relative risk of Alzheimer's and other diseases (see Cordor et al., Scienca 261(1993). Multiplex PCR amplification of SNP loci with subsequent hybridization to ollgonucleotida arreys has been shown to be en accurate end raliable mathod of simultaneously ganotyping at least hundreds of SNPs; sea Wang at al., Science, 280:1077 (1998); see elso Schefer et el., Neture Biotechnology 16:33-39 (1998). The compositions of the present Invention may easily be substituted for the errays of the prior art.

Genarally, a sampla contalning e target analyta (whather for datection of tha targat analyte or scraening for binding partnars of the target analyte) Is added to tha array, under conditions suitable for binding of the target analyte to at least one of the captura probas, i.e. generally physiological conditions. Tha presence or absence of tha targat analyte is than datectad. As will be appreciated by

those in the art, this may be done in e variety of ways, generally through the use of a change in an opticel signal. This change cen occur via many different mechanisms. A few examples include the binding of e dye-tagged enalyte to the bead, the production of a dye species on or near the beads, the destruction of en existing dye species, a chenge in the optical signeture upon enelyte interaction with dye on bead, or eny other optical interrogatable event.

In a preferred embodiment, the change in opticel signal occurs es a result of the binding of a target analyte that is lebeled, either directly or indiractly, with e detectable label, preferably an optical label such as a fluorochrome. Thus, for example, when a proteinaceous target analyte is used, it may be either directly labeled with a fluor, or indirectly, for exemple through the use of e labeled antibody. Similarly, nucleic ecids are easily labeled with fluorochromes, for example during PCR amplification as is known in the art. Alternetively, upon binding of the target sequences, e hybridizetion indicator may be used as the lebel. Hybridization indicators preferentielly essociate with double strandad nucleic ecid, usually reversibly. Hybridizetion indicators include intercalators and minor end/or major groove binding moieties. In e preferred embodiment, intercalators may be used; since intercalation generally only occurs in the presence of double stranded nucleic acid, only in the presence of target hybridizetion will the label light up. Thus, upon binding of the target enelyte to a cepture probe, there is a new optical signal generated at that site, which then may be detected.

Alternatively, in some cases, as discussed above, the target analyte such es en enzyme generates a species thet is either directly or indirectly opticel detectable.

Furthermore, in some embodiments, a change in the optical signature may be the basis of the optical signal. For example, the interaction of some chemical target enalytes with some fluorescent dyes on the beads mey alter the optical signature, thus generating a different optical signal.

As will be epprecieted by those in the art, in some embodiments, the presence or absence of the target enalyte mey be done using changes in other optical or non-optical signals, including, but not limited to, surfece enhanced Reman spectroscopy, surface plasmon resonence, radioectivity, etc.

The essays may be run under a variaty of experimental conditions, as will be appreciated by those in the ert. A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assey, such as protease inhibitors, nuclease inhibitors, enti-microbial agents, etc., may be used. The mixture of components may be added in eny order thet provides for the requisite binding. Various blocking and washing steps may be utilized as is known in the ert.

In eddition, the present Invention provides kits for the reactions of the Invention, comprising components of the essays as outlined herein. In eddition, e variety of other reagents may be included in the essays or the kits. Theee include reagents like ealts, neutral proteina, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reegents that otherwise improve the efficiency of the essay, such es protease inhibitors, nucleese inhibitors, enti-microbiel agents, etc., mey be used. The mixture of components may be added in eny order that provides for the requisite ectivity.

All references cited herein ere incorporated by reference in their entirety.

CLAIMS

Wa claim:

1. A method of datermining tha identification of a nuclaotide at a datection position in a targat sequanca comprising:

a) providing a hybridization complex comprising said target sequenca end a capture proba attachad to a microsphara on a surface of a patterned substrata; and

b) datarmining the nuclaotide at said datection position.

2. A mathod eccording to claim 1 wherein said hybridization complex comprises said captura probe, an edepter probe, end said target sequence.

3. A method eccording to cleim 1 wherein said substrata is a fiber optic bundle.

4. A method eccording to claim 1 wharain said datarmining comprisas:

a) contacting said microsphara with e plurality of detaction probes aach comprising:

i) a unique nucleotida at the readout position; and

ii) a uniqua datectabla label; and

b) detecting e signal from at least one of seid datactable labels to identify the nuclaotide at the datection position.

5. A method according to claim 1 wherein seid targat sequence comprises a first targat domain directly 5' adjacent to said detection position, wherein said hybridizetion complax comprises said target sequence, said capture probe and an extension primer hybridized to said first target domain of seld target sequenca, end said datarmining comprisas:

a) contacting said microsphera with:

i) a polymerase enzyme;

ii) a plurality of NTPs aach comprising a covelently attached detactabla label;

under conditions whereby if ona of said NTPs basepairs with the base et said detection position, said extanslon primar is extanded by said enzyme to incorporete said label; and

c) identifying the basa at said datection position.

6. A method according to claim 7 wherein each NTP comprises a uniqua fluorophore.

7. A method for according to claim 1 wherain said targat saquenca comprises 5' to 3':

e) a first targat domain comprising en ovarlap domain comprising et laast a nucleotide in the datection position; and

b) e sacond target domain contiguous with said datection position;

wherein said hybridization complex comprises:

e) e first probe hybridized to said first target domain; end

b) e second probe hybridized to said second target domain, wherein seid aecond probe comprises:

i) a detection sequence that does not hybridize with seld target sequence; end

ii) e detecteble lebel;

wherein if seid second probe comprises e bese that is perfectly complementary to eaid detection position e cleevage structure is formed;

seld method further comprising:

e) contacting said hybridizetion complex with a cleevage enzyme that will cleave said detection sequence;

d) forming an assey complex with said detection sequence, e capture probe covalently ettached to a microsphere on e surfece of e substrate, and at least one label;

e) detecting the presence or ebsence of said label es an indicetion of the formation of said cleavage structure; end

f) identifying the base et seid detection position.

8. A method of determining the identificetion of a nucleotide at a detection position in e target sequence comprising a first target domein comprising seid detection position end e second target domain edjacent to said detection position, seid method comprising:

a) hybridizing e first ligetion probe to seid first target domein;

b) hybridizing a second ligation probe to seid second target domein, wherein if seid second ligation probe comprises e base that is perfectly complemantary to said detection position a ligation structure is formed;

c) providing a ligation enzyme that will ligate said first end seid second ligetion probes to form e ligated probe;

d) forming en essay complex with said ligated probe, a capture probe covelently attachad to a microsphere on e surface of a substrate, and et least one label;

e) detecting the presence or ebsence of said lebel es an indicetion of the formetion of said ligation structure; and

f) identifying the base et said detection position.

9. A method of sequencing a plurelity of target nucleic acids each comprising a first domain and a edjacant second domain, seid second domain comprising e plurelity of target positions, said method comprising:

a) providing a plurality of hybridization complexes eech comprising a target sequence and a sequencing primer that hybridizes to the first domain of seid target sequence, seld hybridization complexes ettached to e surfece of a substrate;

b) extending each of said primers by the addition of e first nucleotide to the first detection position using a first enzyma to form an extended primer; and

c) detecting the release of pyrophosphate (PPi) to determine the type of said first nucleotide addad onto said primers.

10. A method according to claim 9 wherein said hybridization complexes are attached to microspheres distributed on eald surface.

11. A method of sequencing a target nucleic acid comprising a first domain and an adjacent eecond domain, said second domain comprising e plurality of target positions, said method comprising:

a) providing a hybridization complex comprising said target sequence and a capture probe attached to a microsphere on a eurface of a pettemed substrate; and

b) determining the identity of a plurality of bases at said target positions.

12. A method according to claim 11 wherein said determining comprises:

e) providing a sequencing primer hybridized to said second domein;

b) extending said primer by the addition of e first nucleotide to the first detection position using a first enzyme to form an extended primer;

c) detecting the release of pyrophosphete (PPi) to determine the type of said first nucleotide added onto said primer;

d) extending said primer by the addition of a second nucleotide to the second detection position using said enzyme; end

e) detecting the release of pyrophosphate (PPi) to determine the type of said first nucleotide edded onto said primer.

13. A method according to claim 11 wherein said determining comprises:

e) providing a saquencing primer hybridized to said second domain;

b) extending said primer by the eddition of a first protected nucleotide using e first enzyme to form en extended primer;

c) determining the identification of seid first protected nucleotide;

d) removing the protection group;

e) adding a second protected nucleotide using seid enzyme; end

f) determining the identification of seid eecond protected nucleotide.

14. A kit for nucleic ecid sequencing comprising:

a) a composition comprising:

i) a substrate with a pattemed surface comprising discrete sites; end

ii) a populetion of microspheres distributed on said sites;

wherein said microspheres comprise cepture probes;

b) en extension enzyme; and

c) dNTPs.

15. A kit according to claim 18 further comprising:

d) a second enzyme for the conversion of pyrophosphete (PPi) to ATP; and

e) e third enzyme for the detection of ATP.

16. A method of detecting e target nucleic acid sequence, said method comprising:

e) attaching e first edepter nucleic acid to a first target nucleic acid sequence to form e modified first target nucleic ecid sequence;

b) contacting said modified first target nucleic ecid sequence with an erray comprising:

i) a substrate with a petterned surfece comprising discrete sites; and

ii) a population of microsphares comprising at least a first subpopulation comprising e first captura probe, such that seid first capture probe end said modified first target nucleic ecid sequence form a hybridizetion complex; wherein seid microspheres are distributed on said surface; end

c) detecting tha presence of said modified first target nucleic ecid sequence.

17. The method according to claim 16 further comprising

a) attaching a second edepter nucleic ecid to e second target nucleic ecid sequence to form e modified second target nucleic acid sequence;

b) contacting seid modified second target nucleic ecid sequence with seld errey, wherein said population of microspheres comprises at least e second subpopuletion comprising a second capture probe, such that said second capture probe end seid modified second target nucleic ecid sequence form a hybridization complex; and

c) detecting tha presence of seid modified second target nucleic acid sequence.

18. A method of detecting a target nucleic ecid sequence comprising:

a) hybridizing a first primer to e first portion of a target sequence, wherein said first primer further comprises an edepter sequence;

b) hybridizing a second primer to e second portion of said target sequence;

c) ligeting said first end second primers together to form a modified primer;

d) contacting seid adapter sequenca of said modified primer with an array comprising:

i) a substrate with a surface comprising discrete sites; and

ii) a population of microspheres comprising at leest e first subpopulation comprising a first capture probe, such that said first cepture probe and said modified primer form a hybridizetion complex; wherein said microspheres ere distributed on said surfece; end

e) detecting tha presence of seld modified primer.

19. A method for detecting a first target nucleic acid sequence comprising:

e) hybridizing at least e first primer nucialc ecid to said first terget sequence to form a first

hybridization complex;

b) contacting said first hybridization complex with e first enzyme to form a modified first primer nucleic ecid;

c) disessocieting said first hybridization complex;

d) contacting said modified first primer nucleic ecid with en erray comprising:

i) a substrete with e surface comprising discrete sites; end

ii) e populetion of microspheres comprising at least e first subpopuletion comprising e first capture probe; such thet seid first capture probe end the modified primer form en assay complex; wherain said microspheres are distributed on said surface; end

e) detecting the presence of the modified primer nucleic acid.

20. A method eccording to cleim 19 wherein steps e) through c) ere repeeted prior to step d).

21. A method according to cleim 19 further comprising:

a) hybridizing et least a second primer nucleic acid to e second target sequence that is substantially complementary to said first target sequence to form a second hybridization complex;

b) contacting said second hybridizetion complex with seid first enzyme to form e modified sacond primer nucleic acid;

c) disassocieting said second hybridization complex; end

d) forming e second essay complex comprising seid modified second primer nucleic acid and e second capture probe on e second subpopulation.

22. A method for detacting a target nucleic ecid saquence comprising:

a) hybridizing e first primer to e first target sequence to form a first hybridizetion complex;

b) contacting said first hybridization complex with e first enzyme to extend said first primer to form e first nawly synthesized strend end form a nucleic acid hybrid that comprises an RNA polymerasa promotar;

c) contacting said hybrid with an RNA polymerase that recognizes said RNA polymerase promoter end generates at leest one newly synthesized RNA strand;

d) contacting seid newly synthesized RNA strend with en array comprising:

i) a substrete with e surface comprising discrete sites; end

ii) a population of microspheres comprising et leest a first subpopulation comprising e first capture probe; such thet seld first capture probe and the modified primer form en assay complex; wherein said microspheres ere distributed on said surfece; and

e) detecting the presence of the newly synthesized RNA strend.

23. A kit for the detection of e first target nucleic acid sequence comprising:

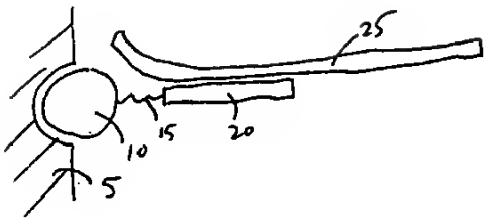
a) at least a first nucleic acid primer substantially complementary to at least e first domain of seid target sequence;

b) at least a first anzyme that will modify said first nuclaic acid primer; and

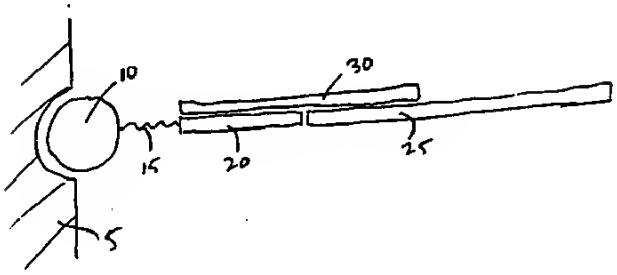
c) an array comprising:

i) a substrata with a pattarned surface comprising discreta sitas; and

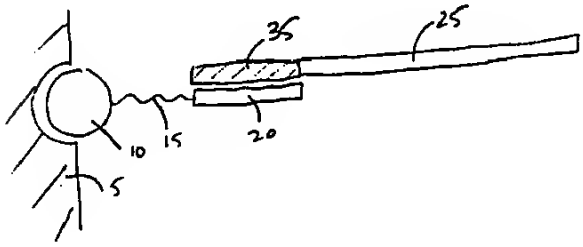
ii) a population of microspheres comprising at least a first and a second subpopulation, wherain aach subpopuiation comprisas a bioactiva agant; wherain said microsphares ara distributad on said surface.



A



B



C

Fig 1

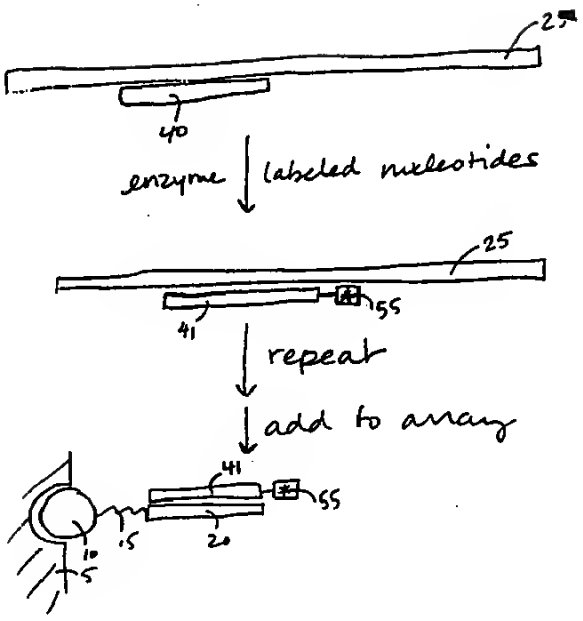
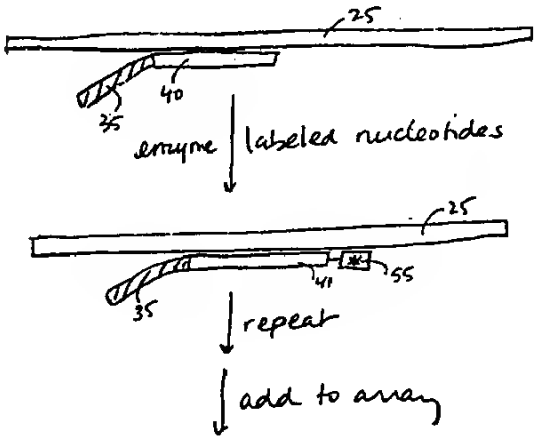

25
40
enzyme
labeled nucleotides
25
55
41
repeat
add to array
41
55
10
15
20
5

Fig 2A


25
40
35
enzyme
labeled nucleotides
25
55
41
35
repeat
add to array

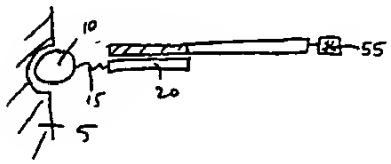

10
55
15
20
5

Fig 2B

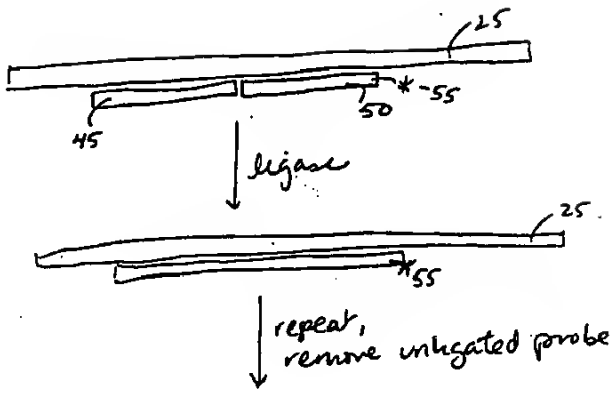

25
*-55
50
45
ligase
25
*55
repeat,
remove unligated probes as needed
10
*-55
15
20
5

Fig 3A

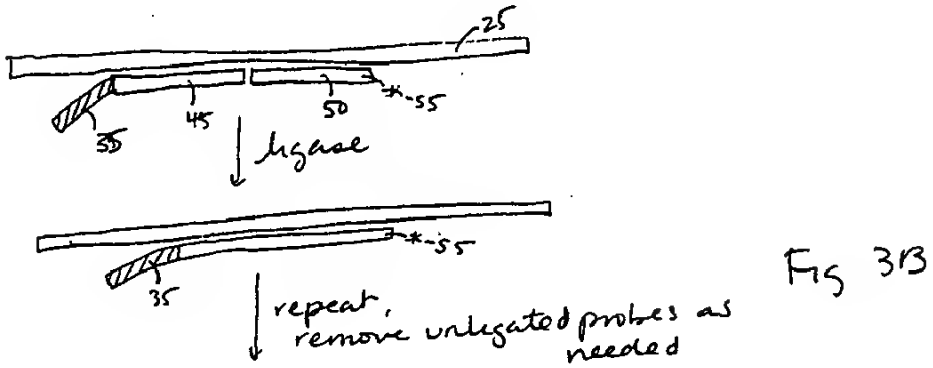

25
*-55
35
45
50
ligase
*-55
35
repeat,
remove unligated probes as needed
35
*-55
10
15
20
5

Fig 3B

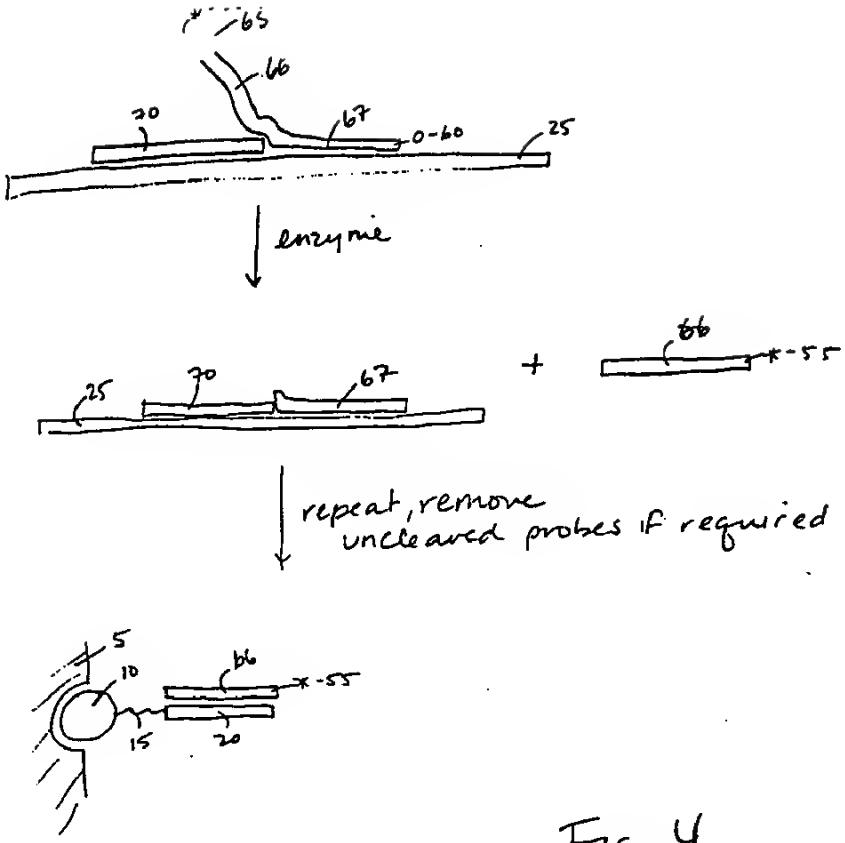

65
66
70
67
60
25
enzyme
66
55
25
70
67
+
repeat, remove uncleaved probes if required
5
10
66
55
15
20

Fig 4

Fig 5A

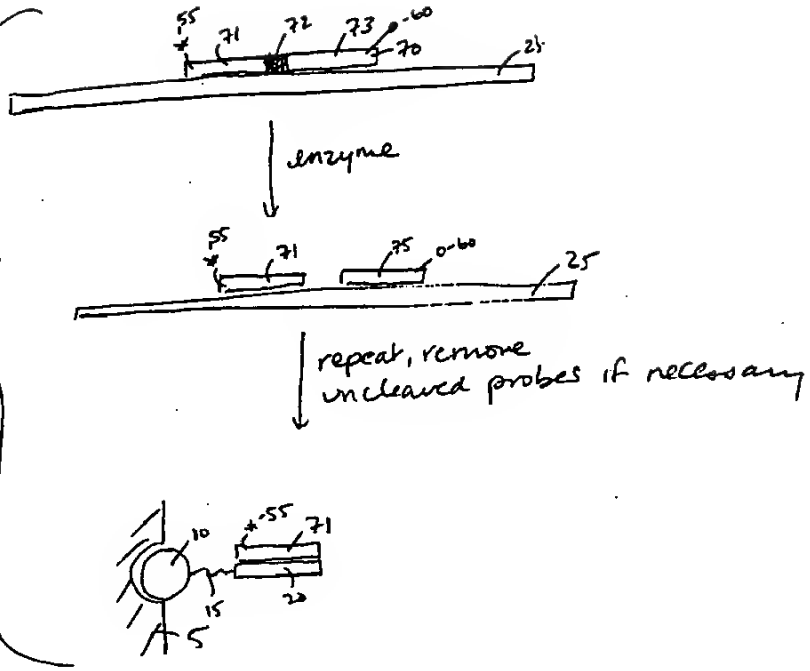
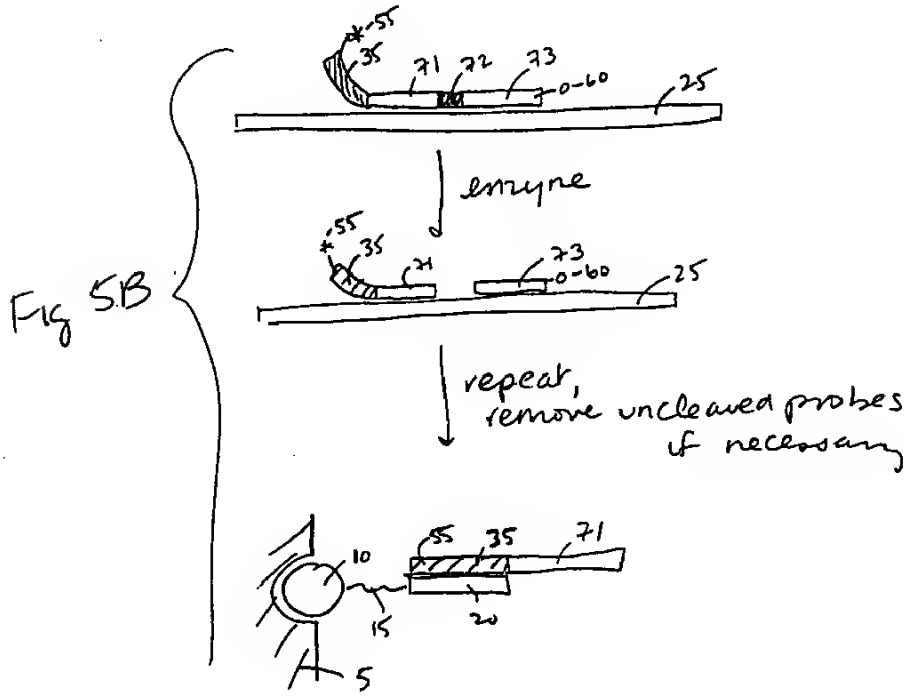

55
71
72
73
60
70
25
enzyme
55
71
75
60
25
repeat, remove uncleaved probes if necessary
10
55
71
15
20
5

Fig 5B


55
35
71
72
73
60
25
enzyme
55
35
71
73
60
25
repeat, remove uncleaved probes if necessary
10
55
35
71
15
20
5

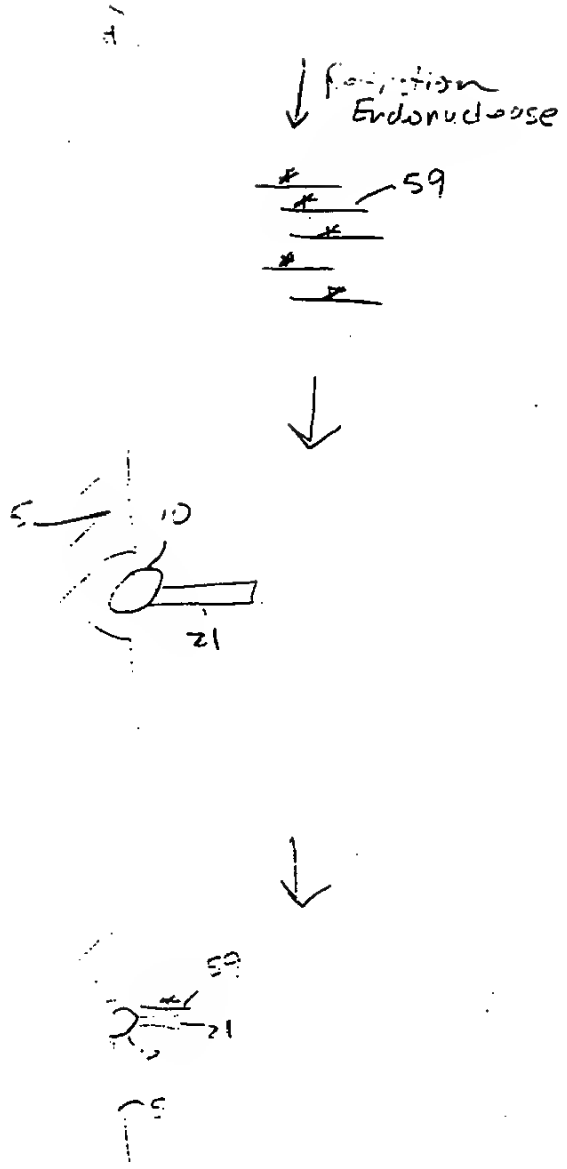
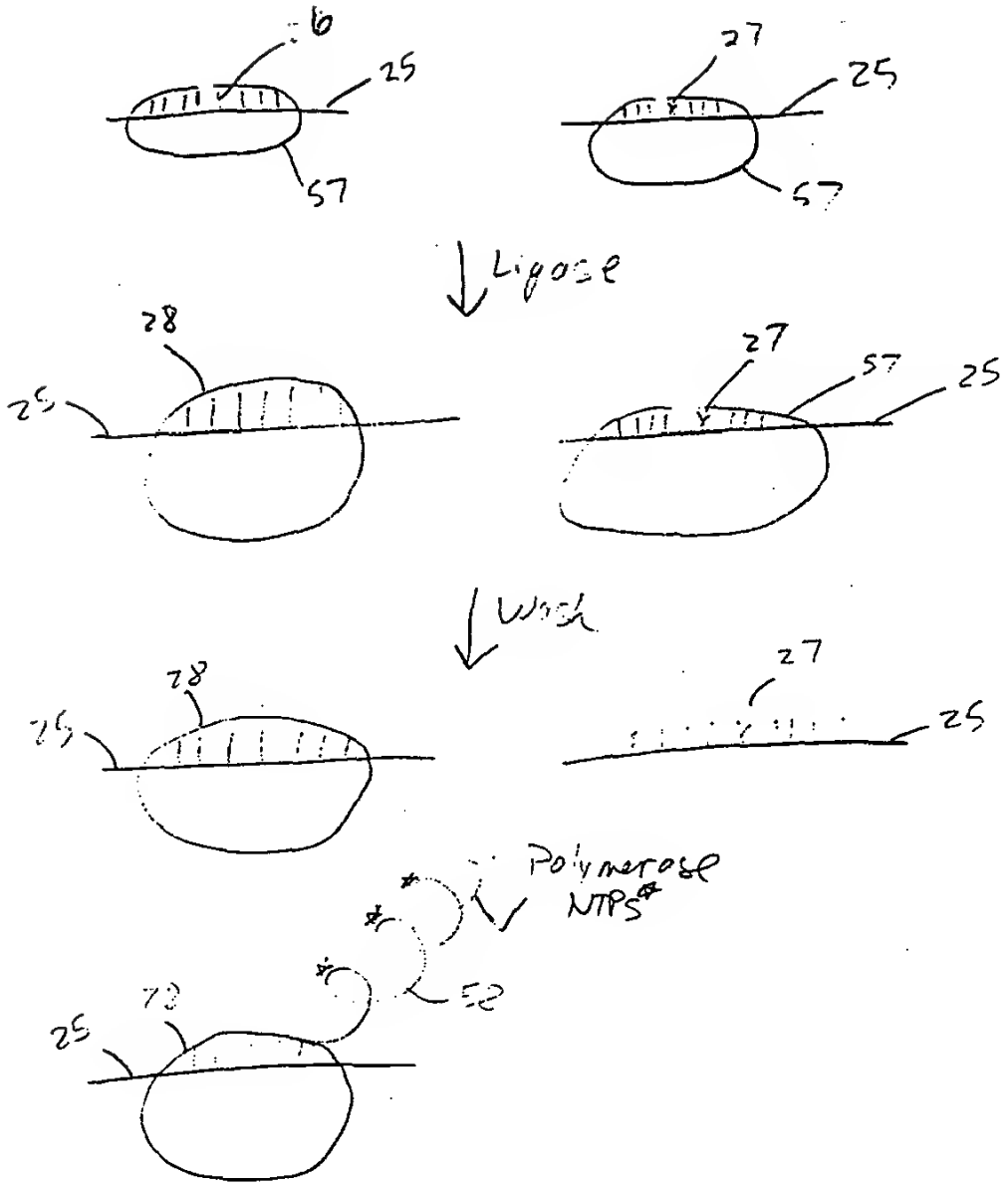
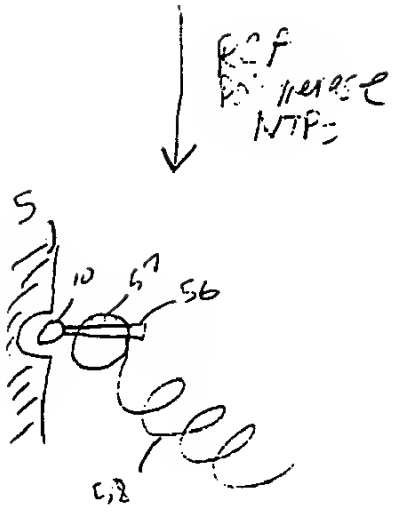


FIGURE 6



7/24

FIGURE 6 (cont)



F

FIGURE 7

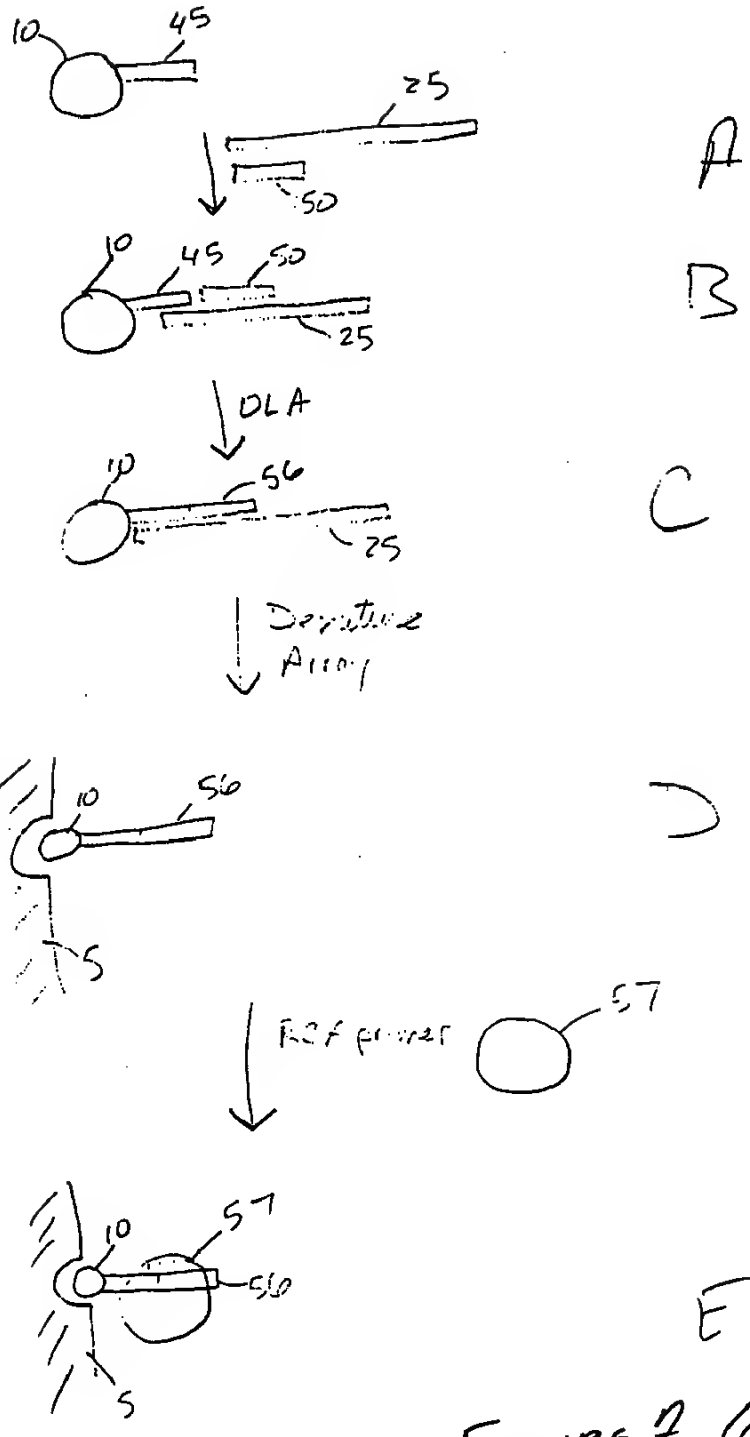


FIGURE 7 (cont)

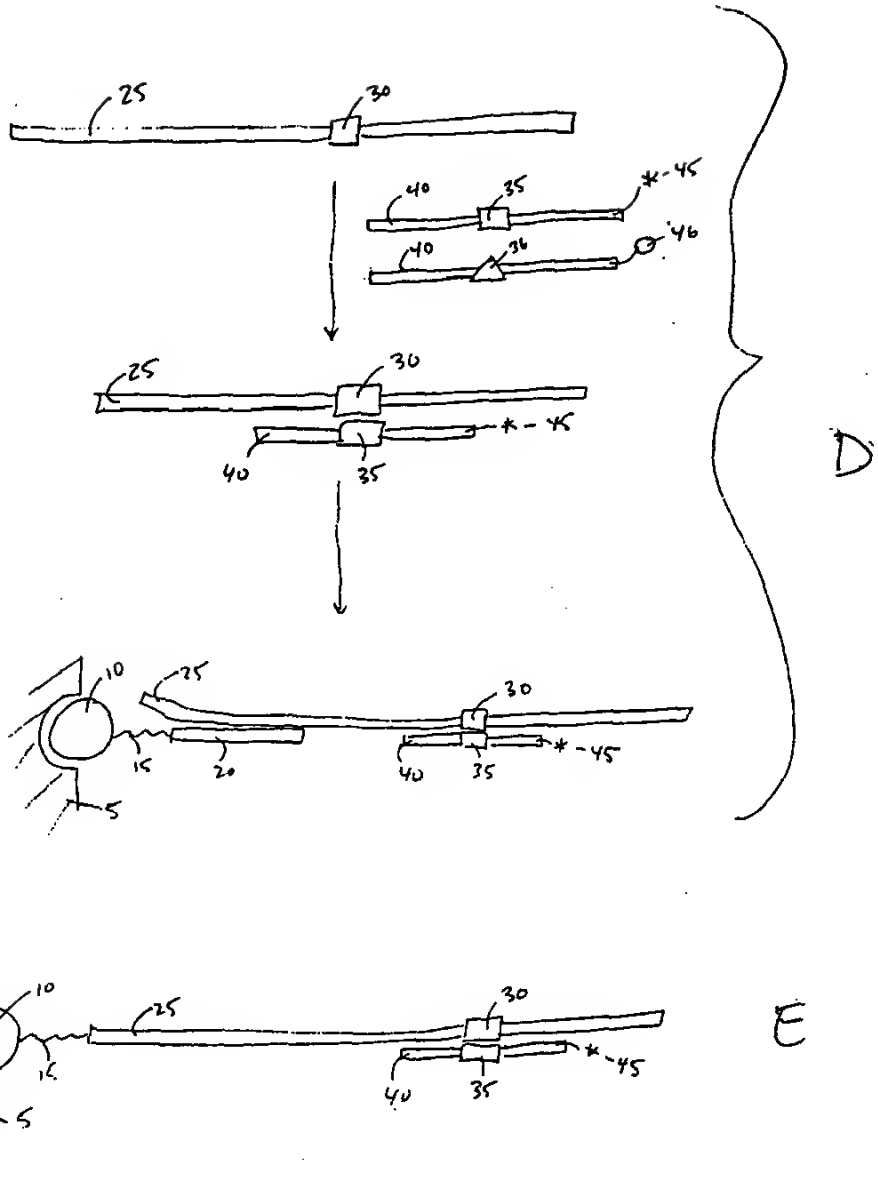
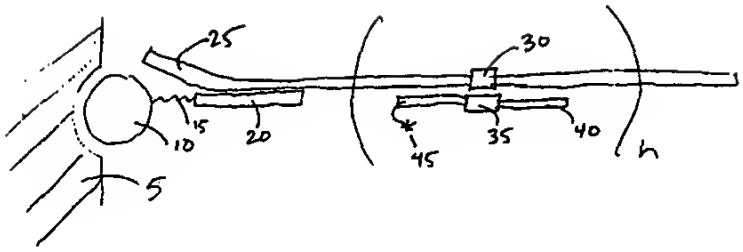
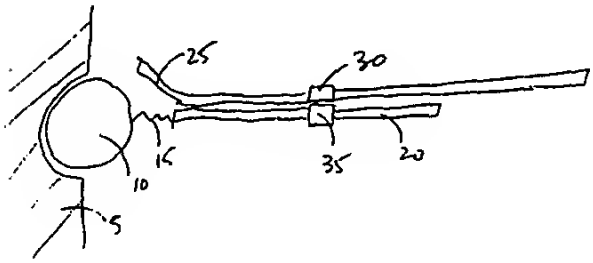

25
30
40
35
45
46
40
36
25
30
45
40
35
D
10
25
30
15
20
40
35
45
5
10
25
30
15
45
40
35
5
E

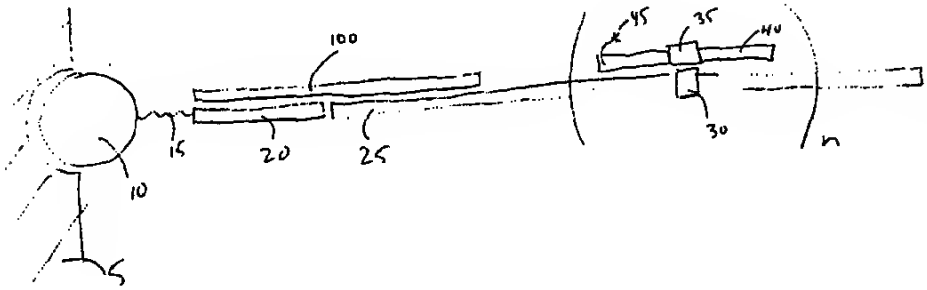
FIGURE 8



A



B



C

FIGURE 8 (cont)

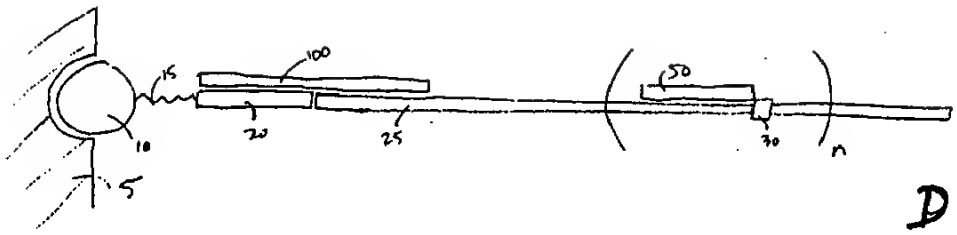
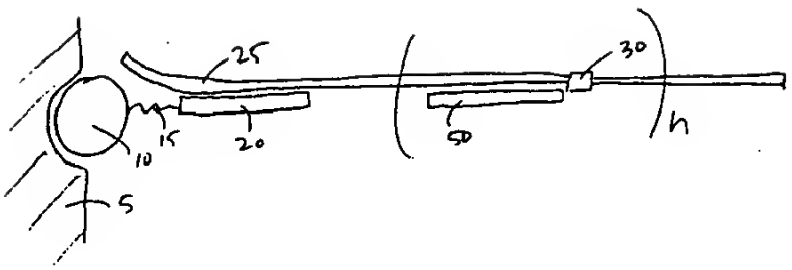
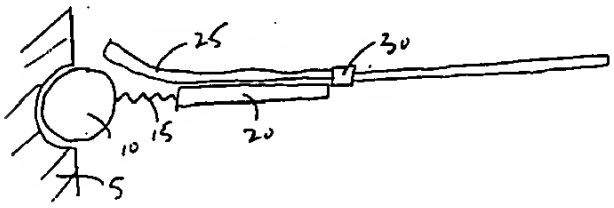


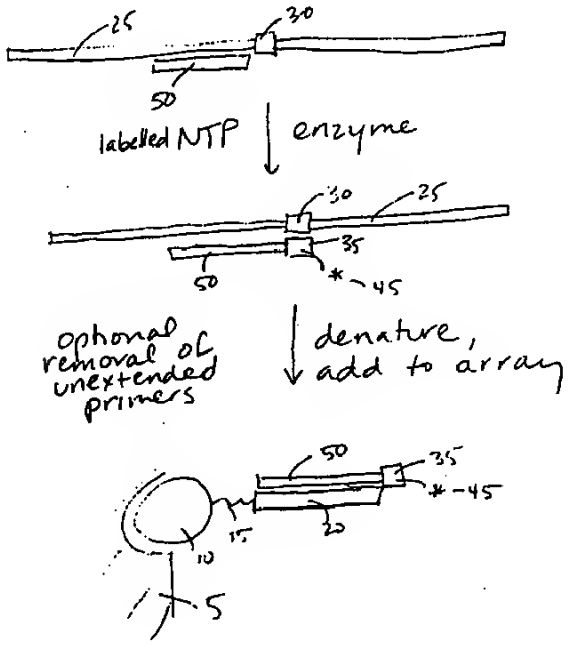
FIGURE 9



A



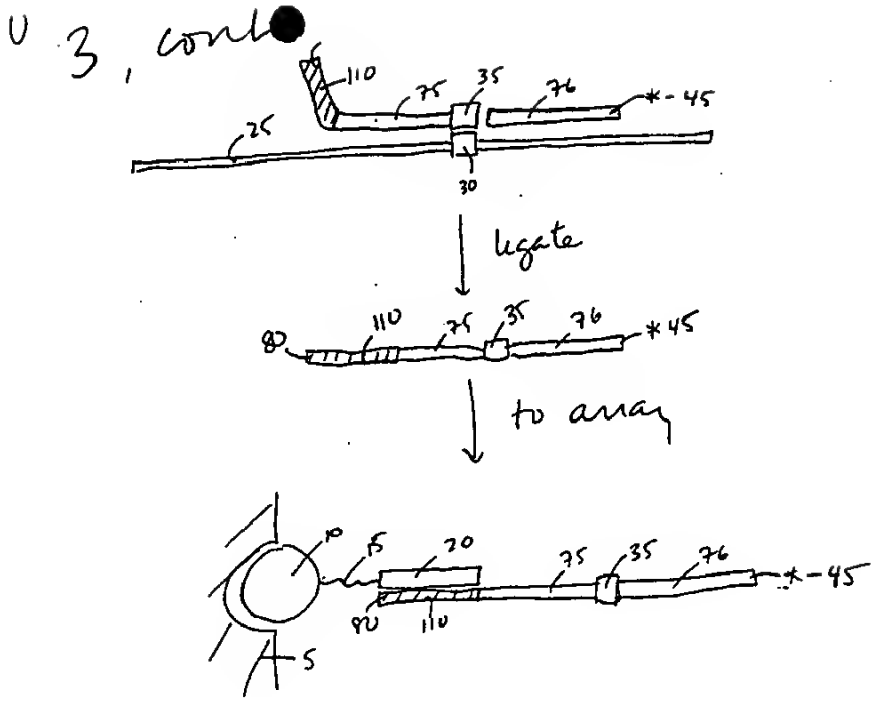
B



C

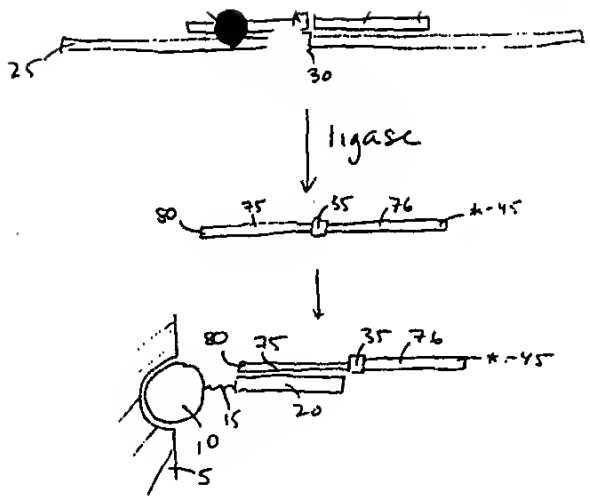
FIGURE 9 (cont)

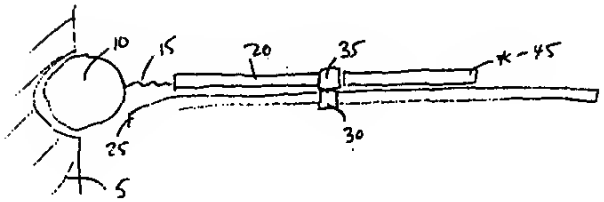
U 3, cont

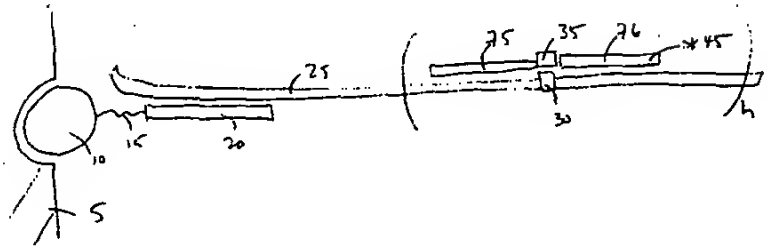


E

FIGURE 10







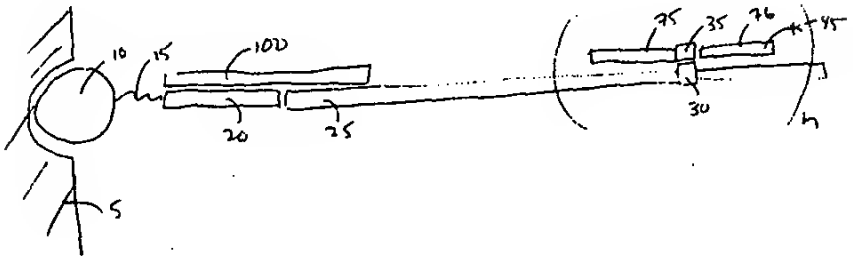


FIGURE 10 (cont)

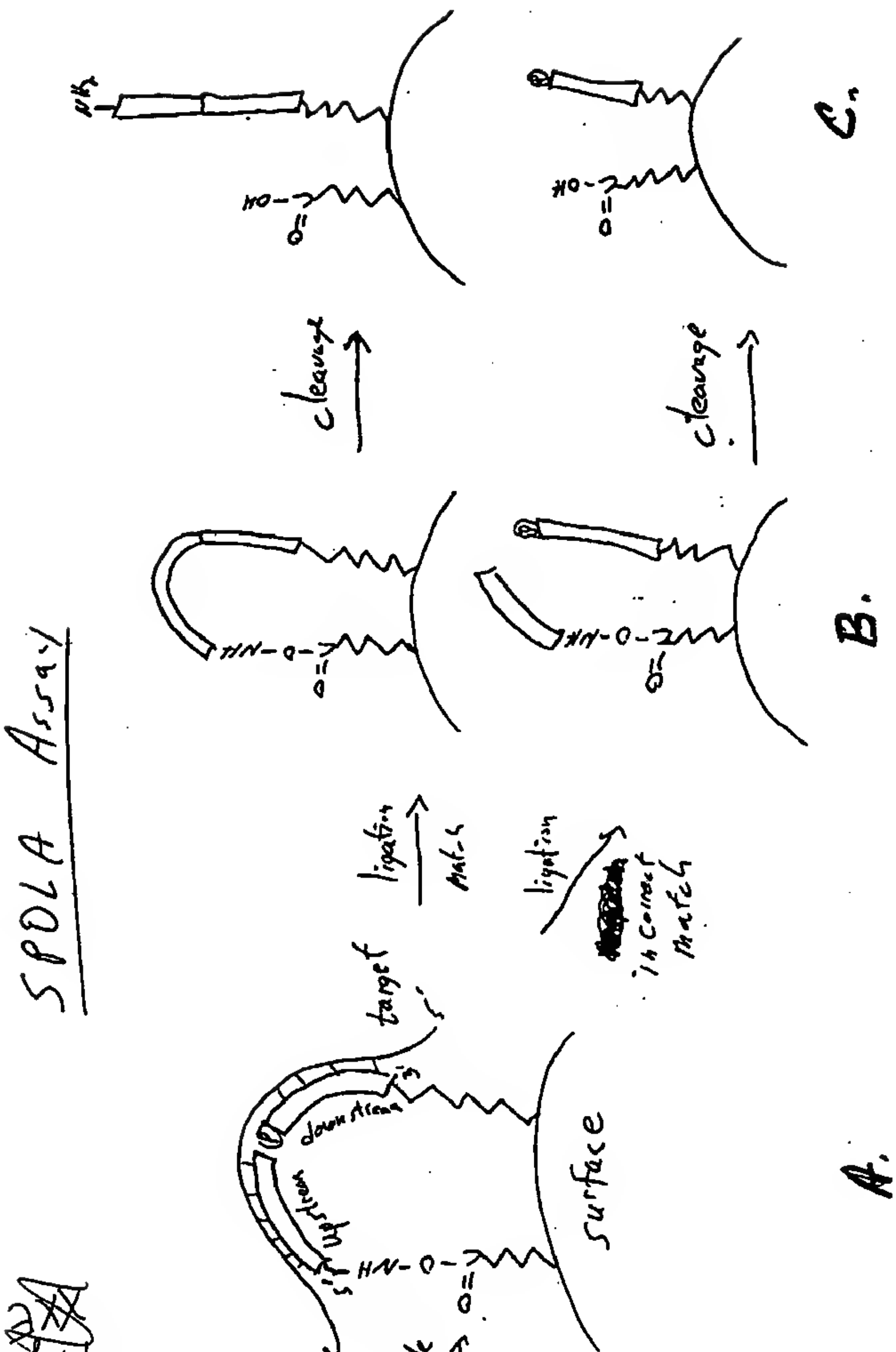

SPOLA Assay
target
5'
3'
downstream
upstream
5'
3'
HN-O-C
O=
surface
cleavable linker
ligation
match
ligation
incorrect match
cleavage
cleavage
C-OH
A.
B.
C.

FIGURE 11

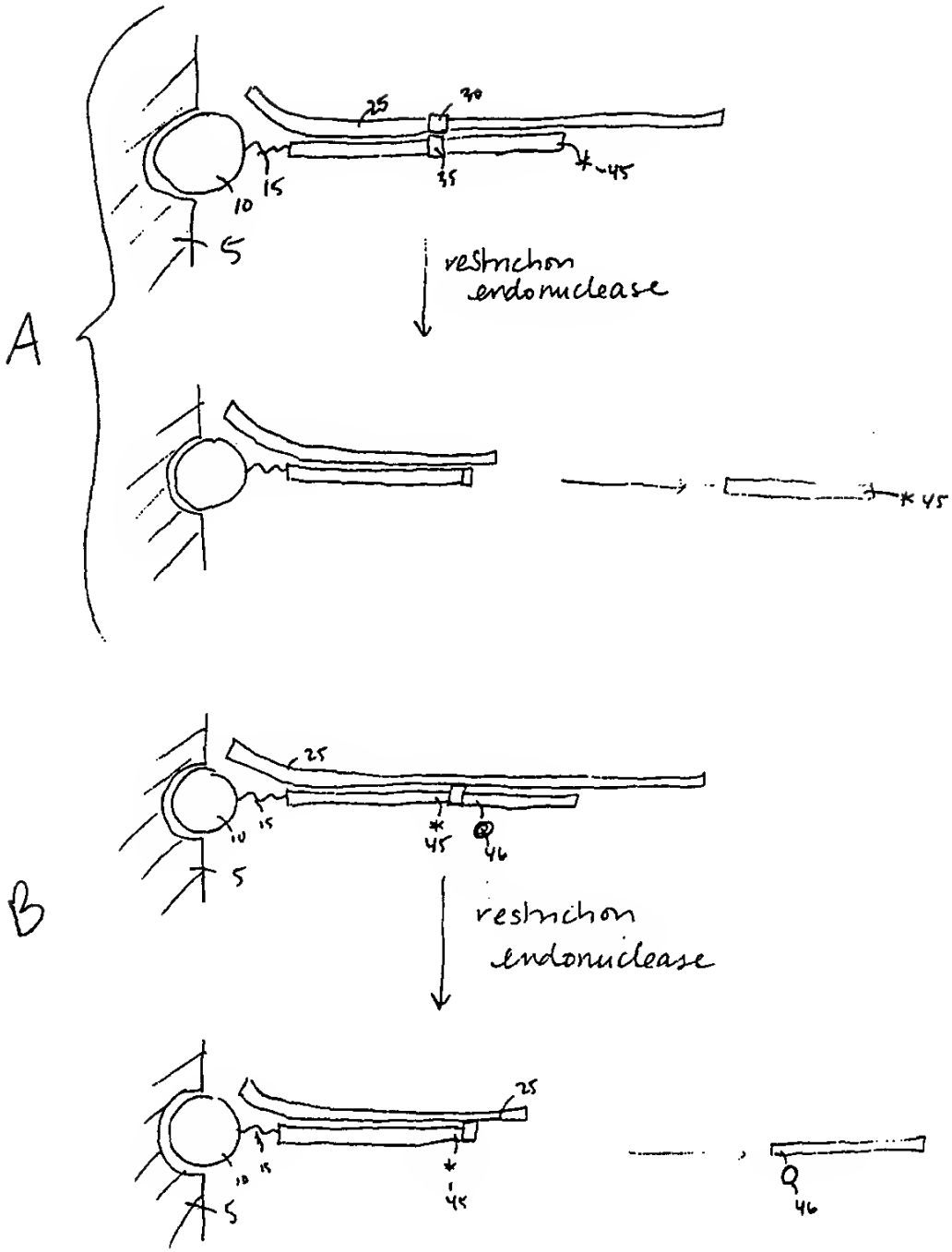
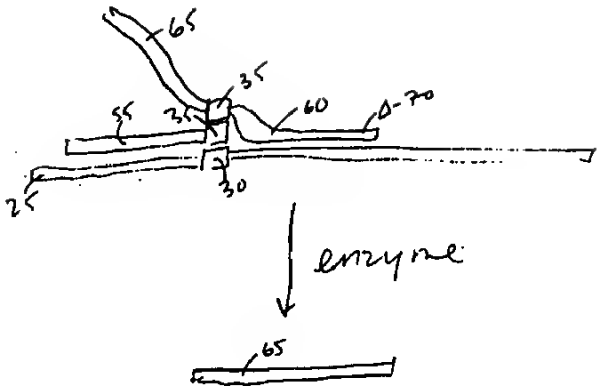

A
25
30
35
45
10
15
5
restriction endonuclease
45
B
25
10
15
45
46
5
restriction endonuclease
25
10
15
45
5
46

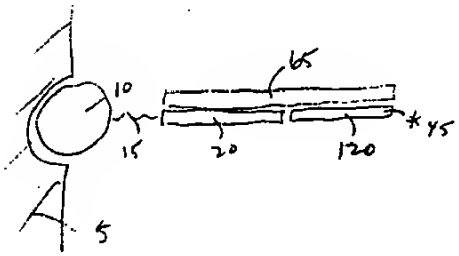
FIGURE 12

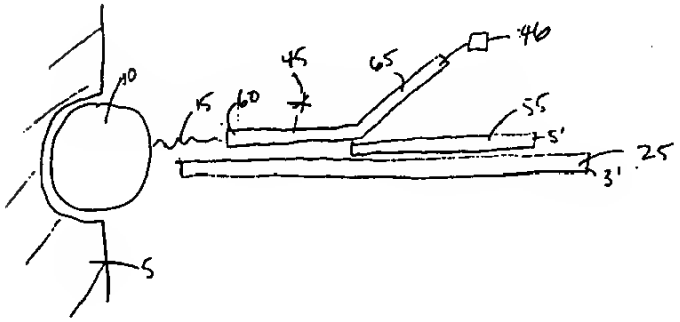


enzyme

optional removal of unreacted primers, optional ligase

D





E

FIGURE 13

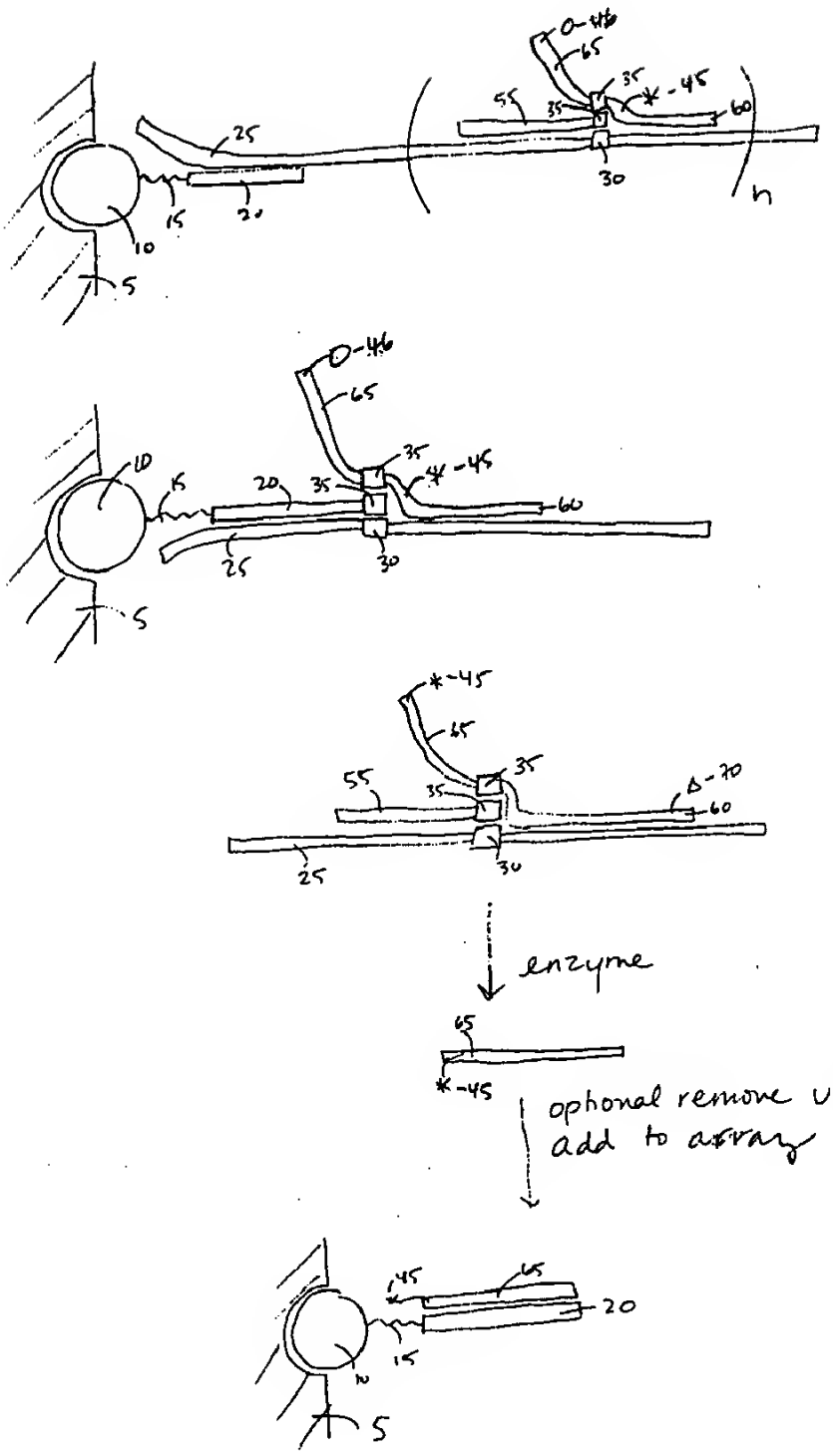


FIGURE 13 (cont)

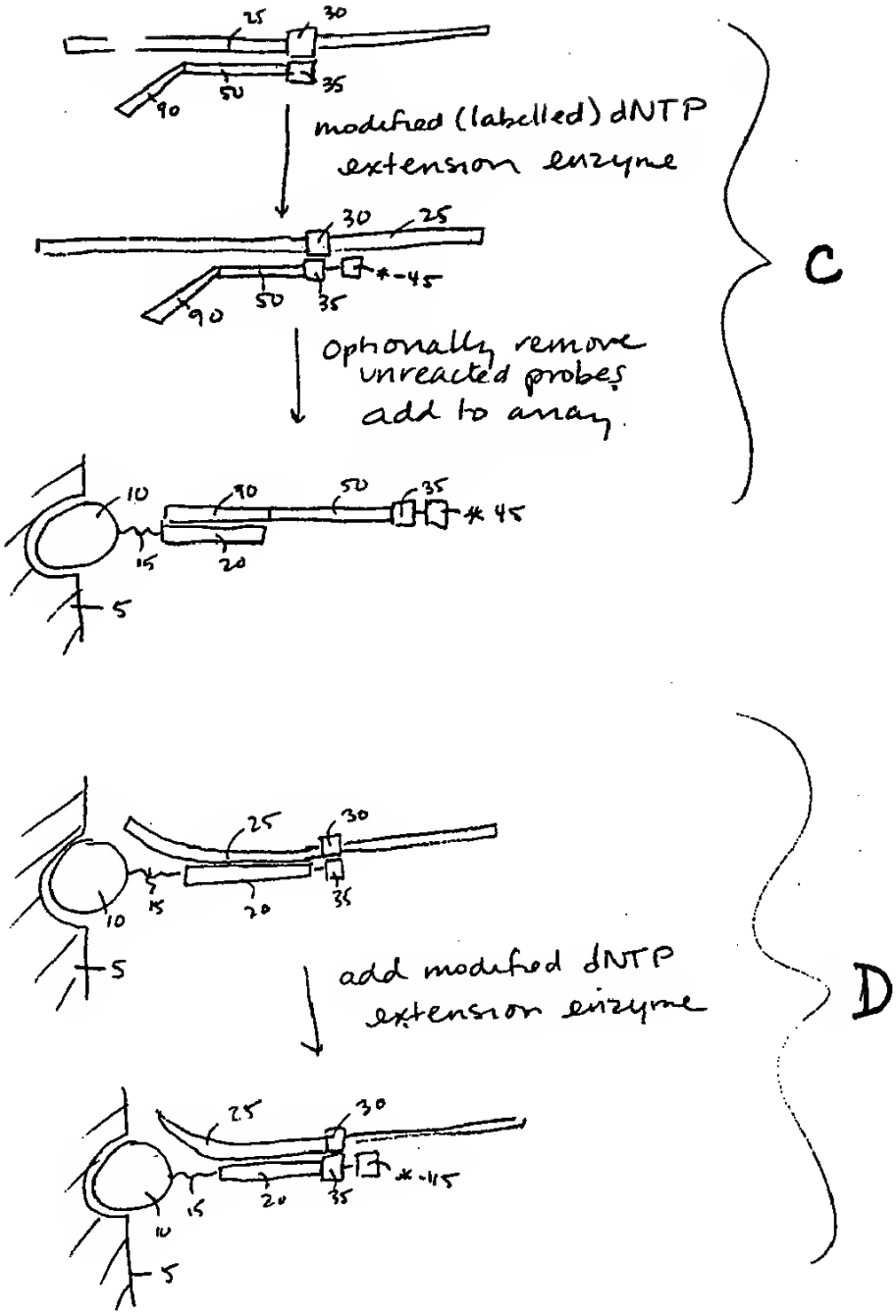
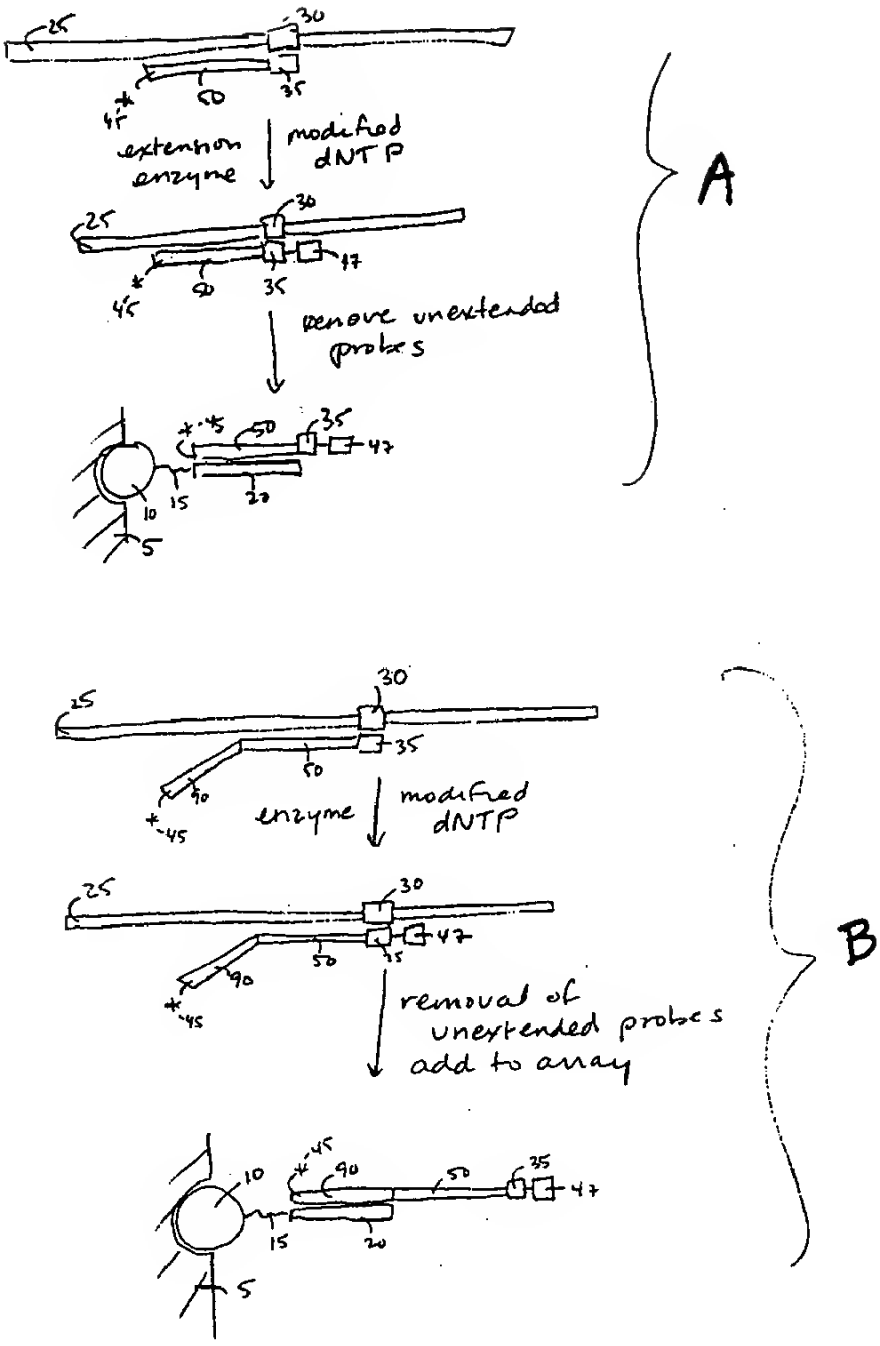
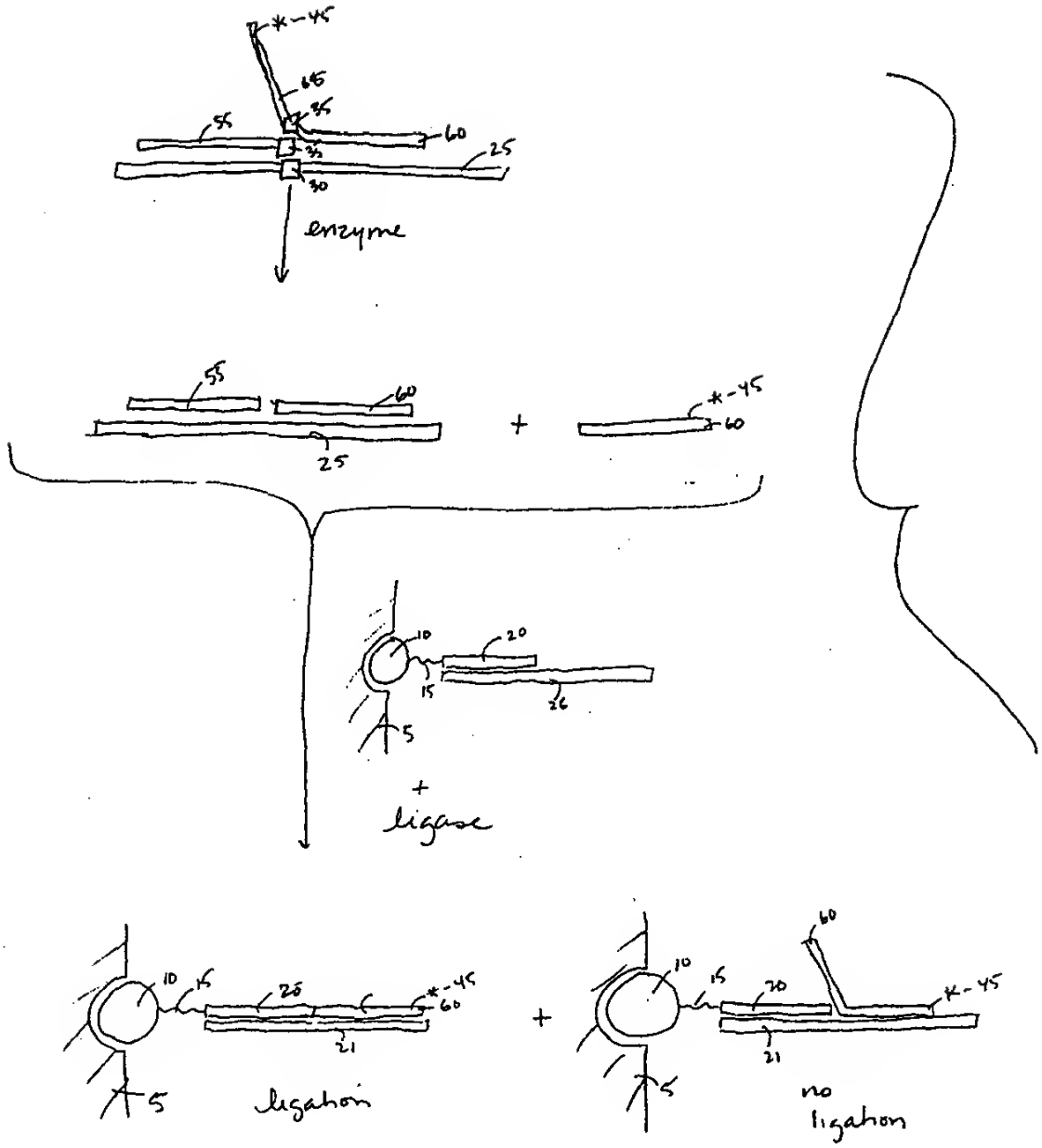

25
30
35
50
90
modified (labelled) dNTP
extension enzyme
30
25
*-45
50
35
90
Optionally remove
unreacted probes
add to array
C
10
90
50
35
*45
15
20
5
25
30
15
20
35
10
5
add modified dNTP
extension enzyme
D
25
30
*-15
15
20
35
10
5

FIGURE 14



2/26

FIGURE 14 (cont)



22/26

FIGURE 15A

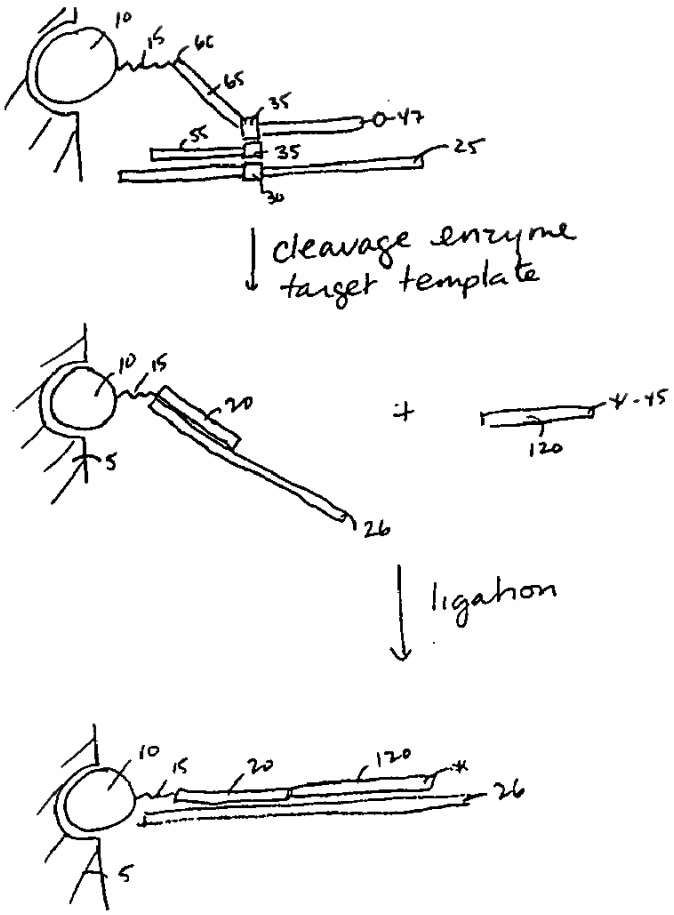

10
15
60
65
35
40-47
55
35
25
30
cleavage enzyme
target template
10
15
20
5
26
+
40-45
120
ligation
10
15
20
120
*
26
5

FIGURE 15B

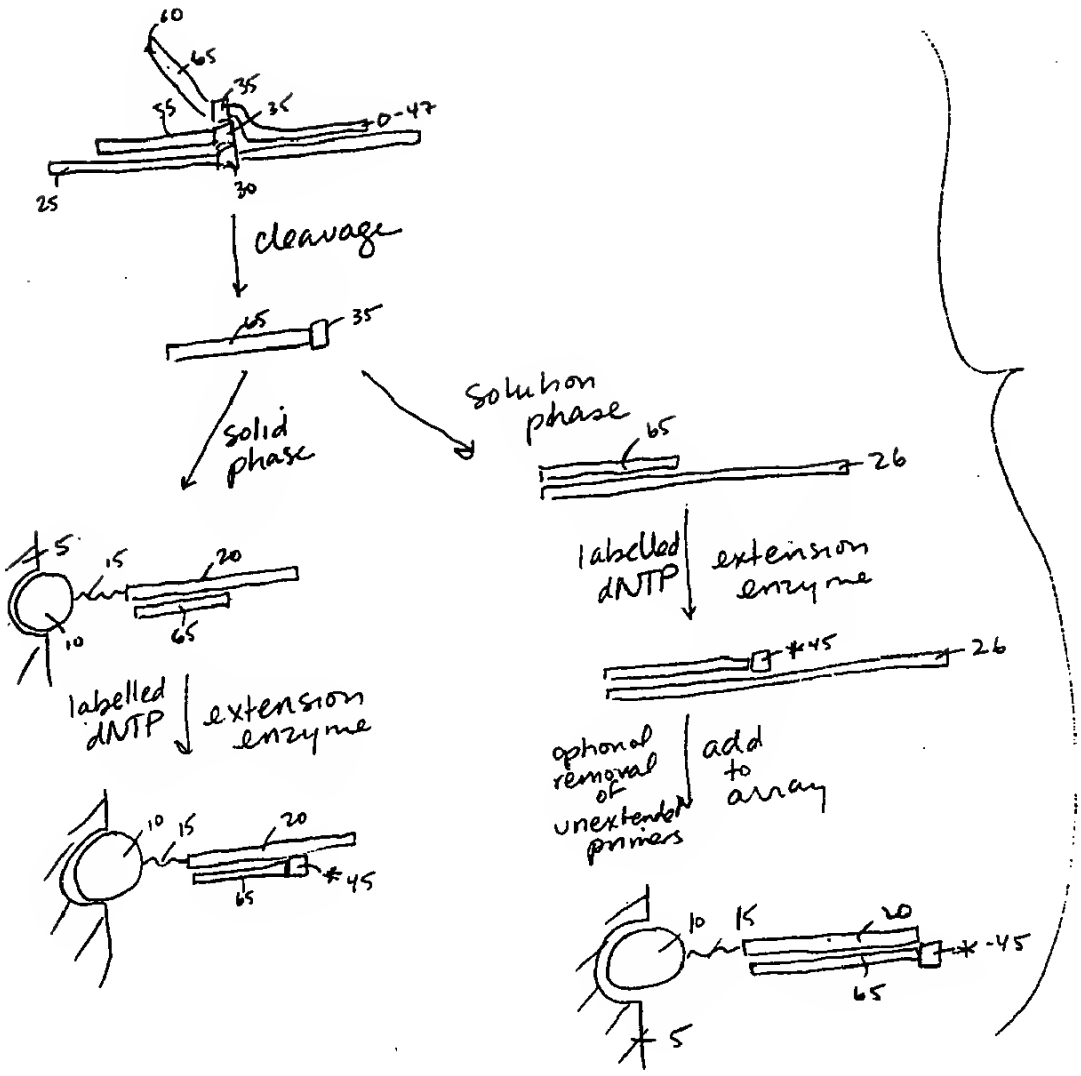

60
65
35
55
35
0-47
25
30
cleavage
65
35
solid phase
solution phase
65
26
5
15
20
10
65
labelled dNTP
extension enzyme
*45
26
optional removal of unextended primers
add to array
10
15
20
*45
65
10
15
20
*45
65
5

FIGURE 16A

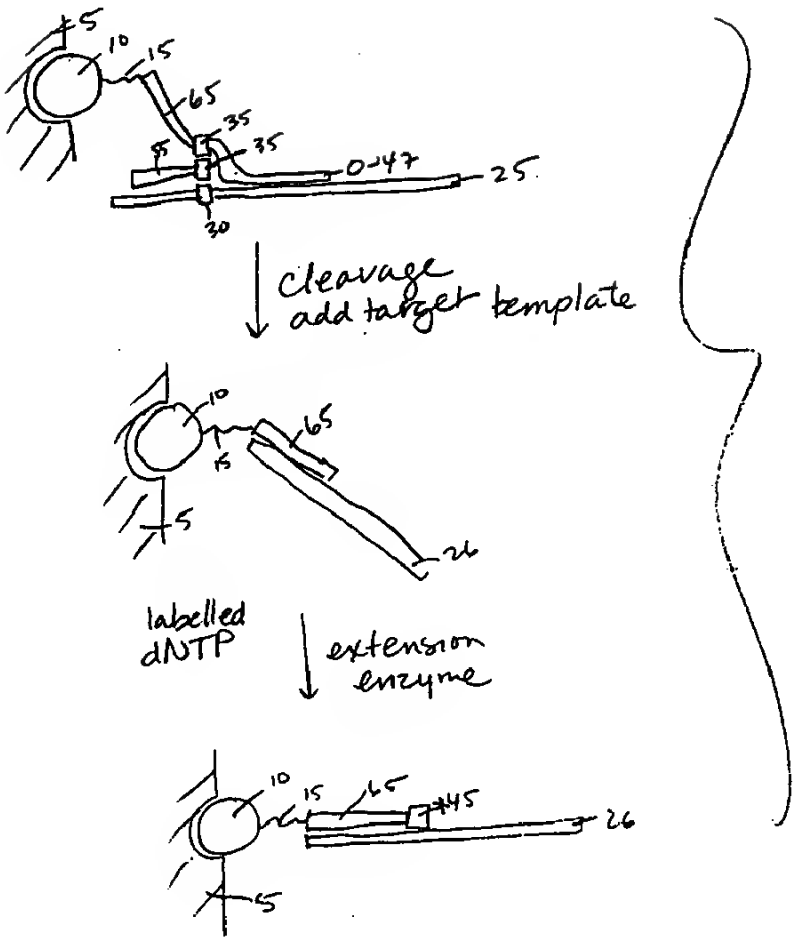
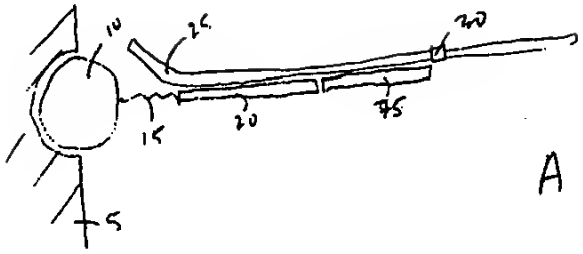
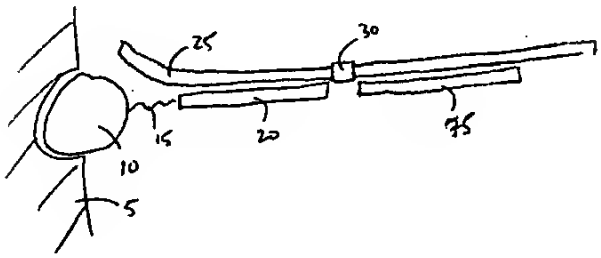

5
10
15
65
35
35
5
0-47
25
30
cleavage
add target template
10
65
15
5
26
labelled
dNTP
extension
enzyme
10
15
65
45
26
5

FIGURE 16B

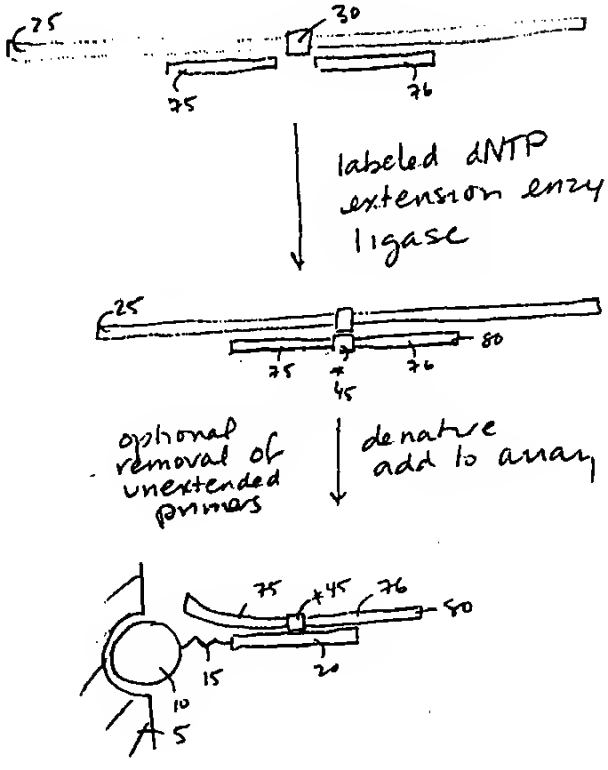

10
25
30
15
20
75
5

A

FIGURE 17


25
30
15
20
75
10
5

B


75
30
75
76
labeled dNTP
extension enzyme
ligase
25
80
75
45
76
optional
removal of
unextended
primers
denature
add to array
75
45
76
80
15
20
10
5